ROBUSTNESS OF AIRNET TO TRAINING LIBRARY FOR SPARSE-VIEW CT

A Dissertation
Presented to
The Academic Faculty

by

Zachary M. Diamond

In Partial Fulfillment
of the Requirements for the Degree
Master of Medical Physics

Georgia Institute of Technology
May 2020

ROBUSTNESS OF AIRNET TO TRAINING LIBRARY FOR SPARSE-VIEW CT

Approved by:

Dr. Hao Gao, Chair
Department of Radiation Oncology
Emory University School of Medicine

Dr. C. K. Chris Wang, Advisor
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Tianye Niu
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: April 21, 2020

“And whoever saves a life …, it is considered as if he saved an entire world.”

– Talmud Balvi: Sanhedrin 37a

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to both Dr. Hao Gao and Gaoyu Chen, for their help and guidance with this work. I would also like to express my sincere thanks to my advisor, Dr. C. K. Chris Wang. His support, advice, and willingness to mentor me throughout my studies at Georgia Tech have been invaluable. I will forever be appreciative.

Additionally, I also thank Dr. Tianye Niu for his support throughout this process and his willingness to serve on my committee. I would also like to give a mention to all of my colleagues, both in NRE and in MP, for their everlasting support, advice, and comradery throughout the past two years.

Lastly, I would like to thank my parents, Scott and Linda Diamond, for their everlasting support, love, advice, and inspiration for me both in my personal life and my academic career.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF SYMBOLS AND ABBREVIATIONS viii

SUMMARY ix

Introduction 1
1.1 Overview 1
1.2 Objectives 2

CHAPTER 2. BACKGROUND 3
2.1 The Physical Basis of X-ray CT 3
2.2 The Technical Basis of X-ray CT 6
2.3 Mathematical Foundations of CT Image Reconstruction 8
2.4 Compressed Sensing 13
2.5 Convolutional Neural Networks & Image Processing 15

CHAPTER 3. Design & Evaluation of AirNET Robustness 19
3.1 AirNET – Background & Underlying Motivation 19
3.2 CT Library Acquisition 22
3.3 Robustness Tests & Evaluation of Robustness Metrics 24
3.3.1 Minimum Mean Square Error 24
3.3.2 Peak Signal-to-Noise Ratio 25
3.3.3 Structural Similarity Index 26
3.4 Image Metric Extraction via TensorFlow 26

CHAPTER 4. Results & Discussions 27
4.1 Image Metric Analysis 27
4.2 Quantitative Image Differences 29

CHAPTER 5. Conclusions & Future Work 33

APPENDIX 35
A) Image Metrics for Robustness Tests 35

B) Time Dependencies for Image Metrics 38

REFERENCES 57

LIST OF TABLES

Table 1	Image Metric Maxima for AirNET (200 Epochs, 48 Layers)	28
Table 2	Image Metric Maxima for AirNET (100 Epochs, 24 Layers, 2 Batches)	29
Tables 3-6	Image Metric Maxima for AirNET (Remaining Cases)	35-37

LIST OF FIGURES

Figure 1	Energy Dependence of X-ray Interaction Probabilities	5
Figure 2	Fan Beam & Parallel Beam Geometry	6
Figure 3	Illustration of X-ray CT Acquisition for a Given Angle	7
Figure 4	2-D Radon Transform of a Brain Phantom	9
Figure 5	Geometry of the Kaczmarz Method	13
Figure 6	Examples of Compressed Sensing Thresholds with Shepp-Logan Phantoms	14
Figure 7	Example of an Artificial Neural Network	15
Figure 8	Schematic of AirNET	22
Figure 9	Geometry of Sparse View Projections	23
Figure 10	Ground Truth, Predicted and Difference of Prostate CT Images	30
Figure 11	Zoomed in View of L-Spine	30
Figure 12	Ground Truth, Predicted and Difference of Lung CT Images	31
Figure 13	Zoomed in View of Femoral Head	31
Figure 14	Ground Truth, Predicted and Difference of Abdominal CT Images	32
Figure 15	Zoomed in View of Right Lung	32
Figures 16-34	Time Dependencies of Image Metrics	38-56

LIST OF SYMBOLS AND ABBREVIATIONS

CT	Computed Tomography
CT-SIM	Computed Tomography – Patient Treatment Simulation
SVD	Single Value Decomposition
TSVD	Total Single Value Decomposition
ART	Algebraic Reconstruction Technique
ReLU	Rectified Linear Unit
MSE	Mean Square Error
SNR	Signal-to-Noise Ratio
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index
MRI	Magnetic Resonance Imaging

SUMMARY

The advances of artificial intelligence and deep learning applied to medical physics are giving rise to numerous applications, ranging from improvements in clinical workflow to the usage of computer-aided diagnosis in preliminary patient screenings. One such advance comes in the form of reconstructing sparsely sampled medical images, whereby a sufficiently trained convolutional neural network would be able to recreate an image.

AirNET is a neural network that reconstructs sparsely sampled CT images by referencing several CT-SIM training libraries for a given case. To test the robustness (the ability of the model to reproduce a correct image given any input) of AirNET, patient libraries of prostate, lung, and abdominal cancers were created, trained, and tested to quantify how well the model accurately predicted the given sparsely-view image. Tests on such networks were performed by running AirNET with different training libraries and different model hyperparameters. Resulting absolute differences between predicted and ground-truth images were taken and shown to be fairly minimal. Additional anatomical images were analyzed on a pixel-by-pixel basis for minute differences in pixel intensities. Image comparison metrics were obtained for each of the tests, as well as their time dependencies. Maxima and minima of such metrics were found to be dependent on both the training library used and the model hyperparameters.

INTRODUCTION

1.1 Overview

In a 2019 study performed by biostatisticians at the University of San Francisco, the overall rate of patients undergoing a diagnostic imaging procedure had increased nearly linearly since the beginning of the 21st century. Of the nearly 136 million imaging examinations reviewed in the study, there had been a roughly twofold increase in the cumulative number of imaging procedures occurring across three different age groups, ranging from younger children to older adults (1). Reasons for such a steady increase in the number of imaging examinations can be attributed to the increased availability of radiological imaging centers, the expansion of government funded healthcare options (Medicaid), and the steady growth of populations in major urban centers. Radiologists and clinicians dealing with imaging-heavy subspecialties have become overwhelmed by such a dramatic increase in treatment workflow, and are therefore requiring to adapt to fast-changing circumstances with respect to obtaining medical images.

As the most common form of medical imaging modality in the United States (2), computed tomography scanners have recently become subject to the "short-time/minimal-dose" constraint in both clinical and industrial settings. An emergency medicine physician specializing in trauma cases, for example, must image a patient and obtain results as quickly as possible in order to ensure his/her survival. In breast tomosynthesis procedures, female patients have reported varying degrees of pain during imaging due to breast compression (3). As a result, computed tomography scanners have been adjusted to take limited numbers of projections, thereby decreasing the total scan time. Even though the scanning time constraint has been met, the overall image quality decreases significantly and any reconstructed images would be unreadable.

The mathematical framework that governs the above problem is not a novel concept. Since the 1970s, the theory of compressed sensing attempts to accurately reconstruct a time-dependent signal from a series of sampled measurements. As developed and eventually proven in several papers from 2004 to 2006, it is mathematically possible to effectively reconstruct a signal given a number of measurements less than a given threshold (4) (5). From this theory, applied mathematicians specializing in image reconstruction algorithms constructed new methods to recreate desired images without much information required. To further improve such techniques, the rise of machine learning, artificial intelligence, and expansive computational frameworks have made the reconstruction of sparsely-viewed tomography images commonplace (6). As such, the medical imaging community has resorted to utilizing such methods to tackle the "short-time/minimal dose" constraint pressed by clinicians and scanner developers.

1.2 Objectives

The goal of this work was to determine the efficacy and robustness of AirNET, an artificial neural network designed to reconstruct and predict computed tomography images that have been sparsely sampled. By utilizing an appropriate selection of patient CT scans from pre-radiation therapy treatment simulation (CT-SIM), pre-processing the scans by sparsely resampling the data, and further separating into different training libraries based on specific tumors, it was possible to train a neural network on a given library and then test the network's ability to reconstruct a sparsely-sampled image from a different library. Tests to determine the robustness of AirNET were performed by training and testing the network with different hyperparameters and evaluating any changes of the resulting image metrics. In addition, several different image evaluation techniques are demonstrated to further display the versatility of AirNET.

CHAPTER 2. BACKGROUND

2.1 The Physical Basis of X-ray CT

From its inception nearly fifty years ago, x-ray CT has revolutionized all aspects of modern medicine. To this day, x-ray CT is the most frequently used imaging modality, with approximately 70 million scans taking place annually in the United States alone (7). Its uses have become more varied in the past several decades, with applications in radiation therapy (simulating and accurately positioning a patient for treatment) and non-proliferation (imaging of shipping vessels for unwelcome radioactive material) to name a few. However, the primary utilization of x-ray CT still remains highest in diagnostic medicine. As a form of indirectly ionizing radiation, the x-rays impart a radiation dose onto the patient, roughly on the order of 25 – 100 mGy (2), which is approximately 100 times the dose from a standard x-ray radiograph. Due to this fact, it was estimated that nearly half of all the radiation dose received in medical procedures each year is due to x-ray CT (7).

In essence, the physical problem of x-ray CT is to determine and evaluate the distribution of linear attenuation coefficients (denoted by μ) from projections of path integrals that are evaluated when x-rays are transmitted through an object. In its most elementary form, the mathematical expression for x-ray attenuation can be given as,

$$I = I_0 e^{-\mu x} \quad (1)$$

where I is the intensity of the incident x-rays, I_0 is the initial x-ray intensity, μ is the linear attenuation coefficient, and x is the object thickness. The linear attenuation coefficient is defined as the probability per unit length (hence common units of cm^{-1} or mm^{-1}) that an incident x-ray

will be attenuated. In clinical practice, however, Equation 1 cannot be utilized in its elementary exponential nature; the distribution of linear attenuation coefficients inside the human body is a series of values and varies as a continuous function. Therefore, it is more fitting to rewrite Equation 1 as the following:

$$I = I_0 e^{-\int \mu(x,y) ds} \quad (2)$$

It is assumed that the distribution of linear attenuation coefficients in the region of interest is entirely spatially dependent, and therefore requires an integration over the total path length of the incident x-ray. Furthermore, it can be assumed that such trajectories are not straight through the medium due to attenuation. In practice, Equation 2 can be discretized into voxels and then rewritten into the following form:

$$-\ln \frac{I}{I_0} = \int \mu(x,y) ds = \sum_i \mu_i x_i y_i \quad (3)$$

Thus, given the initial and transmitted intensities of the x-rays, it is possible to extract the distribution of linear attenuation coefficients, with the caveat that the exact trajectories of the incident x-rays are known.

Since most regions of the body are not uniform and are composed of different materials, it is necessary to return to the definition of the linear attenuation coefficient μ. Because μ represents a cumulative probability per unit length, all possible physical interactions of the incident x-ray must be accounted for. That is, the linear attenuation coefficient can be written as a sum of the probabilities per unit length of photoelectric, Compton, pair production, and other interactions with a given material. In addition, because attenuation is material dependent, it is

common practice to normalize the linear attenuation coefficient by material density, producing a quantity known as the linear mass attenuation coefficient. As each of the physical processes mentioned above is dependent on the energy of the incident x-ray, it is possible to plot the energy dependence of the linear attenuation coefficient for various materials, as shown in the figure below.

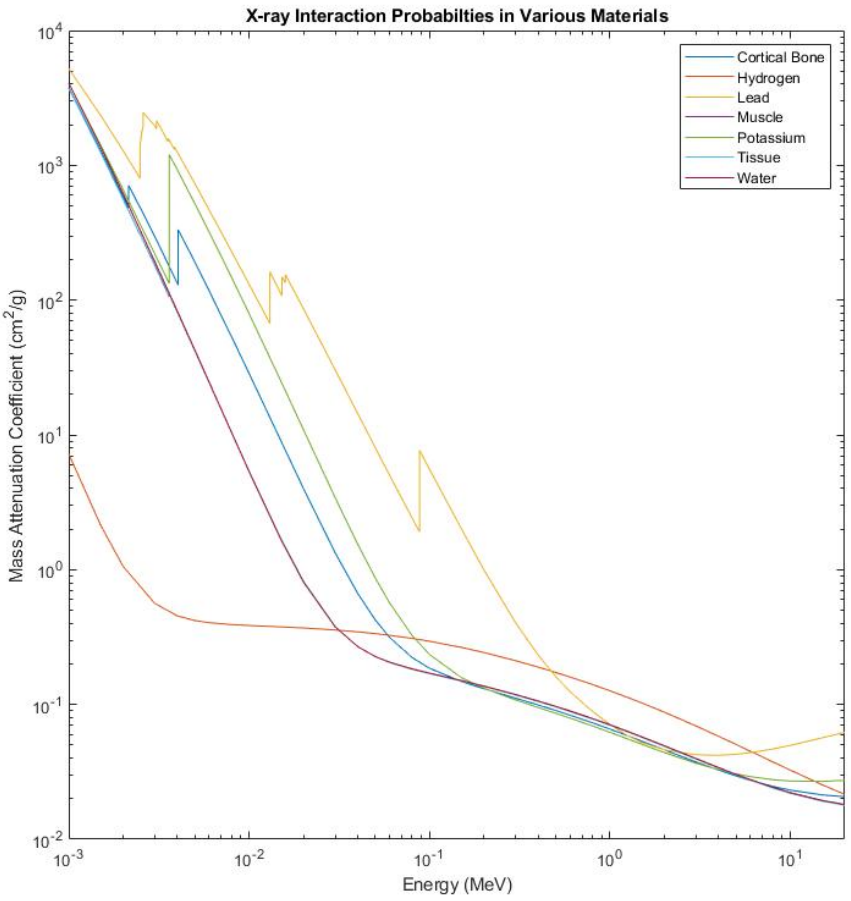


Figure 1 - Energy Dependence of X-ray Interaction Probabilities (NIST)

As a reference, the range of indecent x-ray energies in diagnostic imaging is roughly between 20-300 keV. In terms of physical processes involved, photoelectric absorption and

Compton scattering are the dominant phenomena and therefore contribute most to the value of the linear mass attenuation coefficient. With respect to Compton scattering, is has been shown (8) that the cross section is nearly proportional to the electron density, normally comparable to the value of the proton number Z (or Z_{eff} for compounds). As such, the denser a material is in the body (bone compared to tissue, for example), the higher the probability that an incident x-ray will be attenuated.

2.2 The Technical Basis of X-ray CT

Technological requirements for x-ray CT, in simple terms, are the extension of those required for planar radiography. The x-rays are generated from *bremsstrahlung* radiation, a process in which a stream of thermionic emitted electrons is incident on a thin tungsten target. Emitted x-rays then are able to travel through a patient's body based upon the geometrical configuration of the tomographic imaging system. In the parallel beam configuration, x-rays are uniformly spread out from a planar source as they pass through a patient, whereas in the fan beam configuration, x-rays are emitted from a singular source and spread out radially towards the patient (see figure 2). Recently developed CT imaging systems have vastly utilized fan beam geometries in recent years

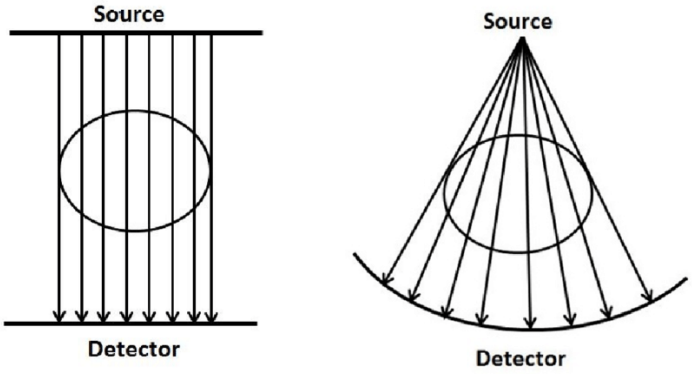


Figure 2 - Parallel Beam & Fan Beam Geometry
(Theeda et al, 2017)

to develop new methods and procedures to accurately image a patient and obtain the best image possible.

Figure 3 depicts how such an imaging system operates at a given projection angle. As an x-ray tube projects x-rays according to one of the above geometries, the resulting x-rays pass through an object across a given path. The attenuated x-rays then interact with a single or series of detectors (commonly inorganic scintillators) which then feed information into the digital acquisition framework that reconstructs the desired image. To describe the specific locations in which the x-rays are attenuated, it is common practice to discretize the imaging volume into a series of voxels,

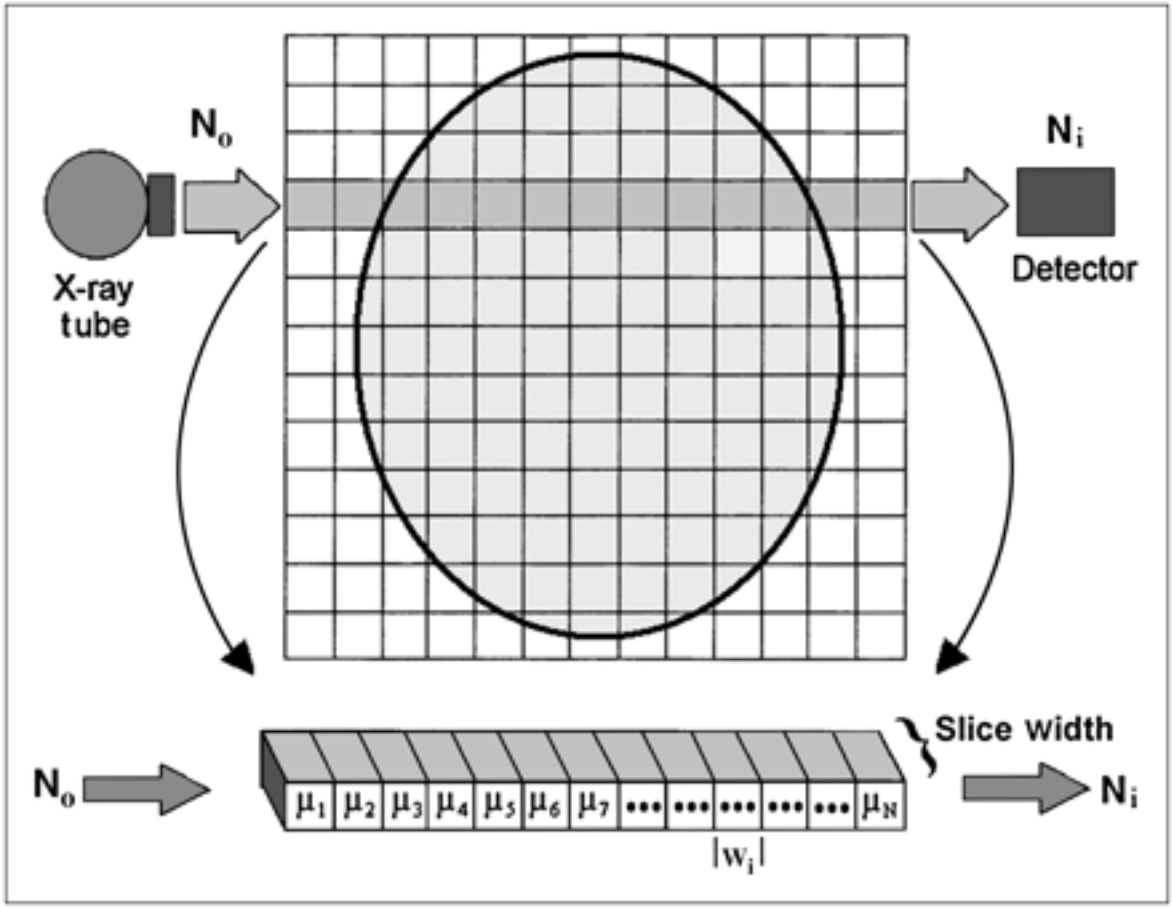


Figure 3 - Illustration of X-ray CT Acquisition for a Given Angle (Goldman, 2007)

and define a system matrix to encapsulate all x-ray path lengths in question. The setup described

in Figure 3 is mostly associated with CT imaging systems in the so-called "third-generation" (9), that is, the x-ray beams utilize a fan beam geometry as the device rotates around the imaging volume. While this may be technically simple, "third-generation" scanners are susceptible to different types of artifacts, especially if a broken detector is present. Moreover, imaging times are fairly long (a few seconds) when compared to more modern machines. Helical CT, in which the images are obtained from the spiral motion of the source/detector setup, allows for faster imaging times. Applied CT imaging systems, such as cone-beam CT (CBCT) and C-arm CT, are designed on similar principles but serve different purposes in the clinic (on-board imaging for cancer treatments and interventional radiology, respectively).

2.3 Mathematical Foundations of CT Image Reconstruction

As stated above, the fundamental mathematical problem in CT image reconstruction is to evaluate the line integrals that depict the x-ray intensity along a given path through an object. For this problem alone, there is extensive research into applied mathematics, machine learning, and optimization theory to develop and test methods in order to obtain the best image possible. In typical practice, there are two main methods of image reconstruction, namely filtered back projection and the iterative method.

Filtered back projection is perhaps the most direct method to obtain the linear attenuation coefficients for each voxel in the imaging volume. In order to obtain this information from the line integrals depicting the x-ray trajectories in the imaging volume, a change of coordinates is required to determine the projection of each x-ray trajectory. Specifically, for parallel beam reconstruction, we can write an equation of a line in the following way (10):

$$L(r, \theta) = \{(x, y) | x \cos\theta + y \sin\theta = r\} \quad (4)$$

In essence, we have simply defined a line in the plane as a function of its radial distance *r* from the origin point and its angular tilt θ from the horizontal axis. From this definition, we can evaluate the line integral of a projection by parametrizing it as follows:

$$p(r,\theta) = \int_{-\infty}^{\infty} f(x(s), y(s)ds \quad (5)$$

where the function *f* represents the unknown image and *x* and *y* are parametrized by:

$$\big(x(s), y(s)\big) = (r\cos\theta - s\sin\theta\,, r\sin\theta + s\cos\theta) \quad (6)$$

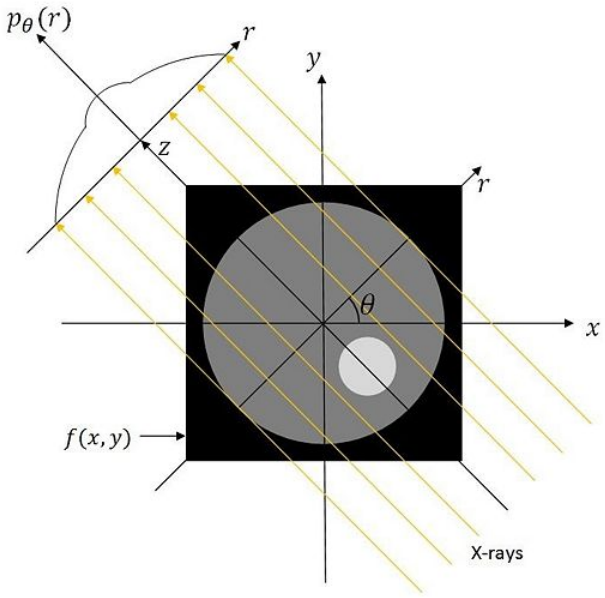


Figure 4 - 2-D Radon Transform of a Brain Phantom (Bouman, 2013)

The function *p* is the projection (of intensities) of the image *f* for a specific angle and has been defined as the 2-D Radon transform of the given image (Figure 4). It is also common practice to rewrite Equation 5 as an integral containing the Dirac delta function, which causes the integrand to be zero elsewhere except on a specific line:

$$p(r,\theta) = \mathcal{R}\{f(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - r)\,dxdy \quad (7)$$

Thus, given a function *f*, one can use Equation 7 to compute the 2-D Radon transform to obtain the image projection. As an example, the circle in Figure (4) has circular symmetry across all projection angles. Therefore, it can be shown that its 2-D radon transform has the following form (assuming that the circle is a unit circle):

$$p(r,\theta) = \begin{cases} 2\sqrt{1-r^2}, |r| \leq 1 \\ 0, elsewhere \end{cases} \quad (8)$$

If we were to take an image of the projection *p*, we would have the sinogram of the image *f*. The sinogram is the fundamental tool for filtered back projection, as it contains all the information required to reconstruct an image for a given angle. In order to now fully reconstruct the image, we utilize the central-slice theorem, which is the basis of the majority of tomographic image reconstruction methods. In terms of operators, the theorem is stated as follows:

$$\mathcal{F}_1\{p(r,\theta)\} = S_1\mathcal{F}_2\{f(x,y)\} \quad (9)$$

As Equation (9) states, the utilization of the Fourier transform allows to extract the image from a given projection. Specifically, the 1-D Fourier transform of a given projection is equal to the slice of the 2-D Fourier transform of the object at the given angle through the origin. By using this theorem, it can be shown that CT image reconstruction occurs in three steps (10) (11): filtering, back projection, and convolution.

Filtering refers to the action of removing low or noisy signals in order to maintain a high image quality resulting from efficient reconstruction. There are many different filters available for

reconstruction, such as the Ram-Lak (ramp) filter (high pass filter), the Hamming filter (window filter), and the Shepp-Logan filter (averaging-ramp filter). This is an essential step to ensure that information is not lost during the reconstruction process, especially with regard to projection sampling. Theoretically, the Nyquist sampling theorem guarantees a lower bound (given by Equation 10) for a signal to be efficiently reconstructed.

$$f_{sample} \geq 2f_{max} \quad (10)$$

Despite this fact holding true, the mathematical theory of compressed sensing does in fact disprove the Nyquist sampling theorem, stating that some under sampled signals can be reconstructed effectively. The mathematical formalities of this theory will be discussed later in this work.

The other method of CT image reconstruction, known as the iterative method, discretizes the calculation of the line integrals that correspond to the x-ray intensity. Referring to Figure 4 above, one can redefine the mathematical problem of determining the linear attenuation coefficients by utilizing the system matrix and x-ray intensity values. By employing language and methods from linear algebra, the linear system of equations governing CT image reconstruction can be written as follows:

$$\boldsymbol{Ax = b} \quad \mathbf{(11)}$$

where A is the system matrix, x is the vector of linear attenuation coefficients, and b is the vector containing the measured x-ray intensities. Despite Equation 11 having a simple algebraic form, the "posedness" of the problem, or the ability for Equation 11 to be solved for x, depends strongly on the features of the system matrix A. Normally, x can be solved for by taking the inverse of the system matrix A and then multiplying by b, the vector containing the x-ray intensities. In

practice, however, A is rarely invertible. A pseudo inverse of A can be determined from the decomposition of the system matrix (SVD, for example), but for the purposes of medical imaging, it is not usually performed in practice extensively due to the presence of noise in the imaging data. To deal with the noise contained in the image, a useful technique is to introduce a regularization term, proportional to the norm of the system matrix. TVSD can be utilized (12) to truncate the decomposition and give the problem a fair algebraic solution, but the quantity of image noise still dominates in this situation.

Despite the fact that A is usually a large matrix (say, with M rows and N columns), it is actually very sparse; not all of the x-ray trajectories pass through every voxel of the system matrix (corresponding to many entries of zero). By utilizing this fact, it is better to describe the problem by calculating the norm of the system matrix and the known parameters. As an example, the problem in Equation 11 can be restated as an optimization problem (13):

$$x = \underset{x}{\operatorname{argmin}} \|Ax - b\|_2^2 + \lambda R(x) \quad (12)$$

where the above expression takes into account the l_2-norm of the known parameters and a regularization term do reduce the amount of image noise. To further discretize the above problem, we can we write Equation 12 as an iterative procedure (14) :

$$x_k = x_{k-1} - \frac{x_{k-1} \cdot (A_k - b_k)}{A_k^2} A_k \quad (13)$$

where A_k is the k^{th} row vector of the system matrix. Equation 13 is also referred to Kaczmarz's method and is a useful result in applied linear algebra. A caveat to this method is that a sufficient initial guess to the problem is required for fast convergence. Equation 13

assumes that for a given number of iterations, the solution will eventually converge with the underlying assumption that the initial guess is sufficient. Geometrically, as shown in Figure 5, the optimal solution is approached after a number of successive orthogonal projections between the two hyperplanes.

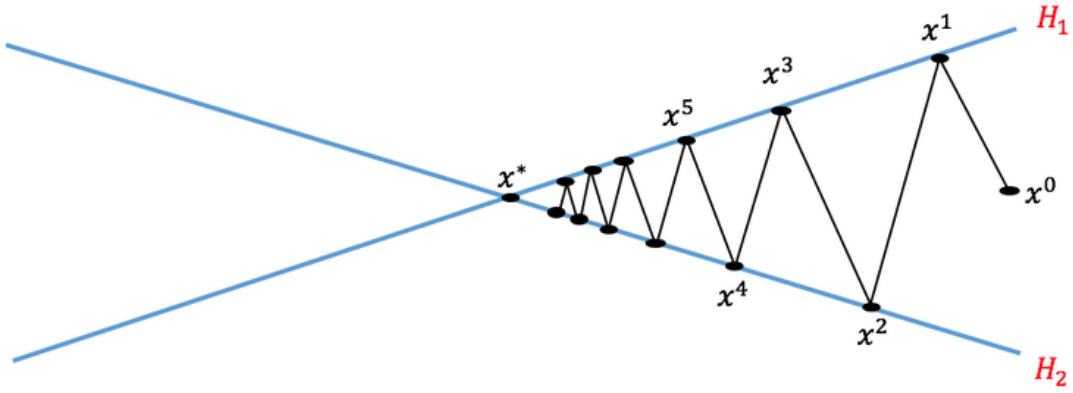


Figure 5 - Geometry of the Kaczmarz Method (Loizou, 2019)

When compared to the filtered back projection method, the computational runtime is much longer due to successive computations (on the order of $O(n^2)$, where n is the number of iterations). However, for simplistic purposes, the iterative method is easier to model due to its discrete mathematical nature. Hybrid CT reconstruction methods, such as combining filtered back projection and iterative methods, are being extensively studied to utilize advantages in both methods. AirNET utilizes one such method, which will be discussed later in this report.

2.4 Compressed Sensing

As stated, prior, the mathematical theory of compressed sensing allows for a stable signal recovery despite imperfect measurements. Although the mathematical formalities are beyond the scope of this report, the sheer importance of this concept and its implications cannot be avoided.

Most of the recent works surrounding compressed sensing by Candès(5) (15) (16), Donoho (17) (18), and others utilize theory from signal processing, linear algebra, and probability theory to determine the efficacy of reconstructing a complete signal with incomplete data. As an example, Candés et al. (15) utilize signal processing theory to compute an upper bound probability that a signal can be reconstructed given a sparse amount of data. As an illustrative example, Candés displays several reconstructed Shepp-Logan phantoms similar to the ones below:

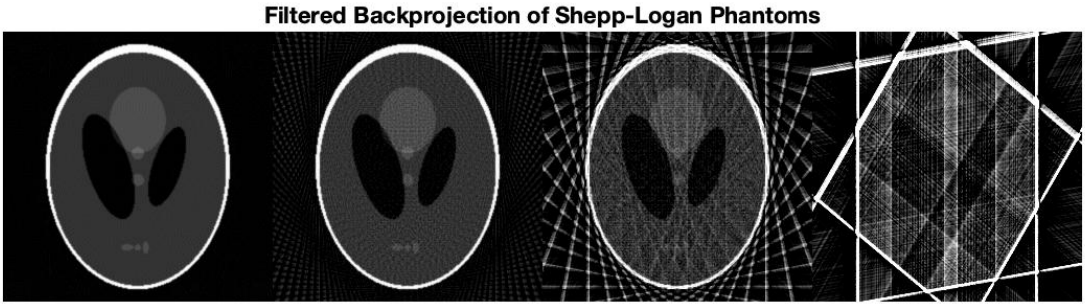


Figure 6 - Examples of Compressed Sensing Thresholds with Shepp-Logan Phantoms (From Left to Right: 180 Views, 60 Views, 24 Views, 4 Views)

Although the best reconstructed image is the furthest most to the left, the reconstructed image with sixty views does have a respectable image quality. That is, most of the main features of the image can still be discernable. As the number of views significantly decreases, however, the image quality worsens significantly, and the reconstruction becomes less effective. A paper by Hanson (19) provides some early work into the theory of optimality of CT image reconstruction algorithms, even providing optimal bounds on image and physical parameters (SNR and absorbed dose, for example). These values are of great importance with respect to sparsely viewed CT image reconstruction, as they provide great insight into the efficacy of a given reconstruction algorithm.

2.5 Convolutional Neural Networks & Image Processing

The rise of statistical machine learning and its applications has become of great use in the field of image reconstruction and processing (20). Specifically, the utilization of artificial neural networks has allowed for significant improvements in classifying images, predicting image features, and overall determination of image quality.

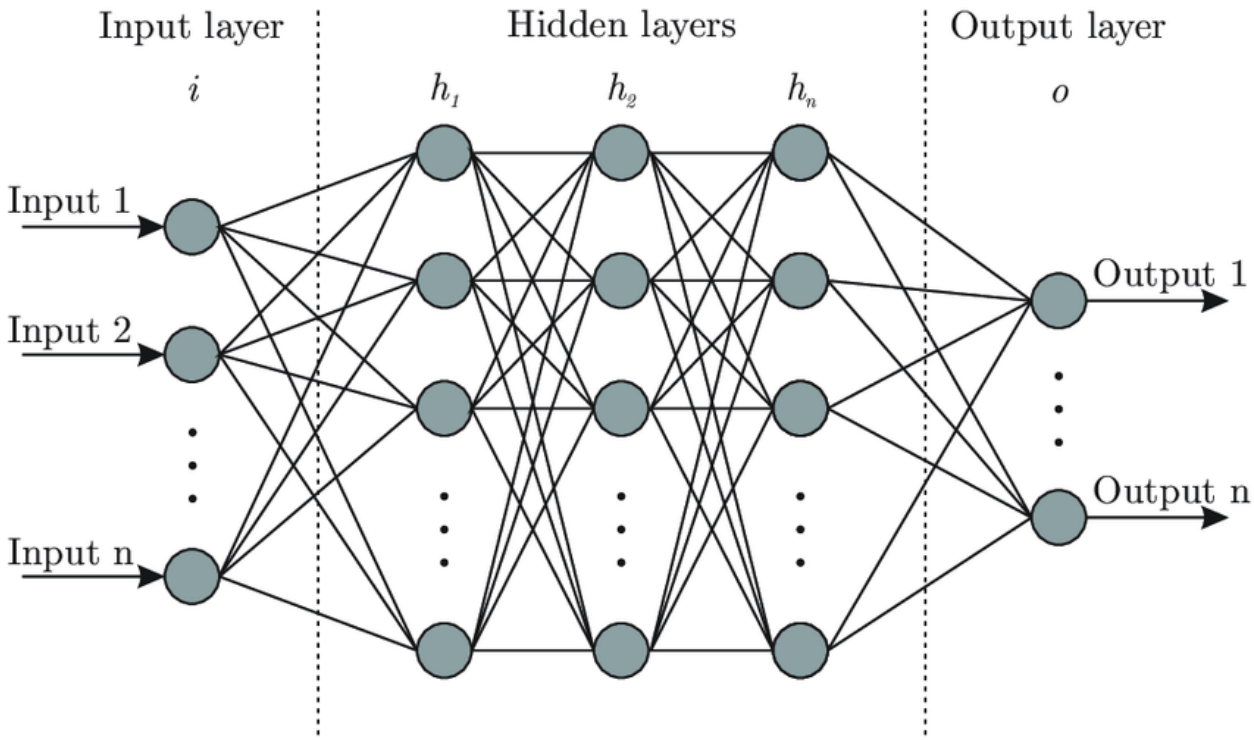


Figure 7 - Example of an Artificial Neural Network (Bre, 2017)

The figure above is a specific example of the architecture of an artificial neural network. Each of the three types of layers, namely the input, hidden, and output layers, are utilized for a different purpose within each network. The input layers, in the case of images, take various parameters or features from the image (pixel intensity, statistical variability, for example) and connect them to the main architecture of the artificial neural network. This architecture, comprised

of numerous hidden layers, is completely dependent on the predictive or computational task at hand. In most tasks regarding image reconstruction or image processing, convolutional neural networks are used the most extensively, since almost all image data follows a "grid-like" topology (21).

The general mathematical problem describing image reconstruction with convolutional neural networks is expansive in nature, but not difficult to understand. We seek to determine a two-variable "learning" function *F(x,v)* that is dependent on the vector containing the training data *v* and the vector *x* containing the model parameters (22). Stochastic gradient descent, a common optimization technique utilized in machine learning, is employed to sufficiently determine the optimal values of the function *F* that minimize a given loss function over successive intervals. Formally, stochastic gradient descent is described by the following expression:

$$x_{k+1} = x_k - s_k \nabla_x L(x_k, v_i) \quad (14)$$

where s_k is the learning rate and the gradient operator acts upon the parameter variables of the loss function. In cases of image reconstruction, the loss function is the mean square error function:

$$L(x) = \frac{1}{N} \sum_{i=1}^{N} \|F(x, v_i) - F_{true}(x, v_i)\|_2^2 \quad (15)$$

The beauty of stochastic gradient descent can be found by choosing the training values *v* at random. From this random selection, it has been shown (21) that with each successive step performed by utilizing Equation 14, the rate of convergence toward an optimal value in the initial iterations is very large. This phenomenon, called semi-convergence, states that initial iterations of

stochastic gradient descent converge faster than the standard gradient descent method. This is a very useful tactic for deep learning methods, especially in convolutional neural networks, due to the fact that the overall time it takes to train a model is often very long. Initial convergence, in the short term is optimal, however as time increases significantly, the convergence behaves in an oscillatory matter.

With respect to updating the data in each iteration within a single epoch (one pass through the training data), it is common practice to apply an activation function to the updated data to eliminate any irrelevant hidden nodes, speed up the training process, and further ensure effective gradient propagation throughout the neural network. Two common activation functions utilized in neural networks are the sigmoidal function (logistic growth function) and the rectified linear unit (ReLU) function:

$$\mathrm{ReLU}(x) = \max(x, 0) \quad (16)$$

The plot of the ReLU function is comparable to that of a ramp filter seen in filtered back-projection. We seek to eliminate irrelevant training data and only keep the important features selected by the neural network, henceforth the reason why the ReLU function is commonly used to achieve such a goal. In image processing, for example, it has been shown that the ReLU activator function is very useful to distinguish small differences in adjacent image pixel intensities (24). Furthermore, the ReLU activator function is used to account for noise in the data, an important feature in image processing that was overlooked with other types of activation functions.

Another useful tactic involved in the training of convolutional neural networks is the technique of batch normalization. Conceived in a 2015 paper (25), batch normalization attempts to perform several tasks, such as estimating successive steps of stochastic gradient descent and

minimizing computational efficiency. Most importantly, batch normalization improves upon a phenomenon called internal covariant shift. After each iteration, the distribution of the nodes of a neural network change. In return, the computational effort to detect such a change is greatly increased, as well as the computation effort required to calculate the associated gradients (26). By normalizing the training data by the mean and variance of a subset of the training data, it becomes easier for the stochastic gradient descent at a given iteration to be calculated, leading to a faster convergence. In practice, the batch normalization factor can be given by the following expression (25) (27):

$$\hat{x}_k = \frac{x_k - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (17)$$

where μ_B is the mean of the minibatch, σ_B is the standard deviation of the minibatch, and ε is a stability parameter. Despite the usefulness of batch normalization, the number of batches utilized might possibly harm the output of the model rather than solidify it. In other words, by reducing the amount of required image features, it may be possible to exclude some data that might have been overlooked in the name of efficient model training. A major trade-off in utilizing batch renormalization in image reconstruction and processing comes in the form of rapid training. Because images are inherently very detailed and include a significant amount of data, batch normalization should be used with caution with images, as important features that could be missed within the training data set will have more severe implications after the training concludes. Evidence supporting this claim is presented and discussed later within this report.

CHAPTER 3. DESIGN & EVALUATION OF AIRNET ROBUSTNESS

3.1 AirNET – Background & Underlying Motivation

AirNET is the convolutional deep learning neural network that will be fully analyzed in this work. To understand the basis of AirNET, it is necessary to break the network down into its two components, namely the reconstruction step and the processing step.

The reconstruction method utilized in this work combines the two most common reconstruction techniques in CT imaging, name the analytical (filtered back-projection) method and the iterative method. Referring back to Equation 12, we wish to determine the optimal values of the linear attenuation coefficients from the measured intensities of the transmitted x-rays that pass through the imaging volume. As such, one can modify Equation 12 to account for both the analytical and iterative reconstruction techniques in the following manner (28):

$$x = \underset{x}{\operatorname{argmin}} \|P(Ax - b)\|_2^2 + \lambda R(x) \quad (18)$$

The operator *P* is defined to be a filtration operator, allowing for analytical reconstruction to be utilized in the reconstruction problem. It can be shown in the Fourier domain that the filtration operator *P* is diagonalizable, but the details involved are beyond the scope of this report. To fully optimize Equation 18, the technique of proximal forward-backward splitting is employed, due to the convex function in the problem as well as the possible non-differentiable regularization term. When that technique is employed, it is possible to split the problem into two parts, namely the data fidelity update step and the image regularization update step (28). To do this, we can utilize the proximal operator (29) (30):

$$\text{prox}_g(y) = \underset{x}{\text{argmin}} \frac{1}{2}\|x - y\|_2^2 + g(x) \text{ (19)}$$

and minimize Equation 18 using the two-part splitting technique (and moving subscripts up for convenience):

$$x^{n+\frac{1}{2}} = x^n - s\nabla f(x^n) \text{ (20)}$$

$$x^n = \text{prox}_{sg}\left(x^{n+\frac{1}{2}}\right) \text{ (21)}$$

where *s* is a parameter that controls the convergence of the algorithm. Equation 20 can be rewritten in a few different ways with respect to the reconstruction technique at hand, namely the iterative technique and the AIR technique. Specifically, for the iterative technique, Equation 20 can be rewritten as (and changing indices):

$$x^{n-\frac{1}{2}} = x^{n-1} - sA^T(Ax^{n-1} - y) \text{ (22a)}$$

where A^T is the transpose of the system matrix (see Equation 11). For the AIR technique:

$$x^{n-\frac{1}{2}} = x^{n-1} - sF(Ax^{n-1} - y) \text{ (22b)}$$

where *F* is an operator that corresponds to the combined filtered back-projection operator and the iterative operator.

To introduce the convolutional neural network into the image reconstruction, Equation 21 can be rewritten to include the directed-convolutional neural network by substituting Equation 22b and expanding out the proximal operator:

$$x^n = x^{n-\frac{1}{2}} + DNN\left(x^{n-\frac{1}{2}}, \theta^{n-\frac{1}{2}}\right) \quad (23)$$

where θ represents the set of parameters of the neural network. Therefore, AirNET can be summarized by combining Equations 22b and 23 as one process (31):

$$\begin{cases} x^{n-\frac{1}{2}} = x^{n-1} - sF(Ax^{n-1} - y) \\ x^n = x^{n-\frac{1}{2}} + DNN\left(x^{n-\frac{1}{2}}, \theta^{n-\frac{1}{2}}\right) \end{cases} \quad (24)$$

The indices above indicate that this method takes intermediate steps into consideration during reconstruction and data processing through the neural network. As such, this "unrolling" feature of AirNET allows to concurrently extract reconstructed image features and processes them throughout the network, while at the same time making the most of the sparsely sampled CT image data.

In the form of a block diagram, Figure 8 (31) describes the operational features of AirNET. Two important features of the network are signified by the green and orange arrows, specifically. The green arrows represent dense connectivity within AirNET, signifying that all extracted data is accounted for with respect to the image reconstruction. The orange arrow represents the dependence on residual learning and stochastic gradient descent, so that the model eventually converges to an appropriate solution.

Because of the dense connectivity and dependence of past iterations, it is intuitive to rewrite the network update step Equation 24 in the following manner:

$$x^n = x^{n-\frac{1}{2}} + DNN\left(\left[x^{n-\frac{1}{2}}, x^{n-\frac{3}{2}}, \dots, x^{\frac{1}{2}}\right], \theta^{n-\frac{1}{2}}\right) \quad (25)$$

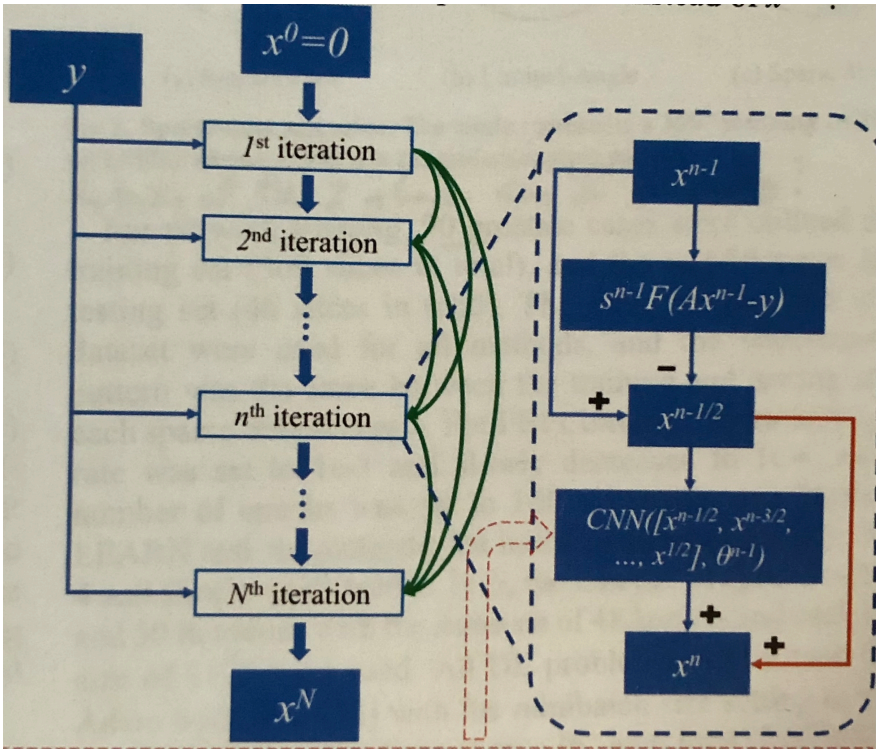


Figure 8 - Schematic of AirNET (Chen et al. 2019)

Since every iteration in the model is accounted for within the model, the data eventually is processed through two convolution and ReLU layers, and then convolved further into a single output layer. The impact of increasing the total number of iterations will be covered later in this report.

3.2 CT Library Acquisition

To create several CT image libraries to be processed through AirNET, CT-SIM scans of approximately 100 cancer patients were downloaded and immediately double-blinded to ensure patient confidentiality. The scans were then split into three libraries based upon cancer diagnoses, namely prostate cancers (approximately 50 patients), lung cancers (approximately 30 patients), and abdominal cancers (pelvis, stomach, etc; approximately 20 patients). The data was then pre-

processed from full-view (360 views, 1° separation) to sparse-view (60 views, 6° separation), and then designating 20% of the training data as testing data. The geometry of the sparse view was chosen for its simplicity and optimality and is most like the original acquisition data. Alternative sparse-data scenarios, such as the limited angle view and the sparse group view, are summarized geometrically in the figure below (31):

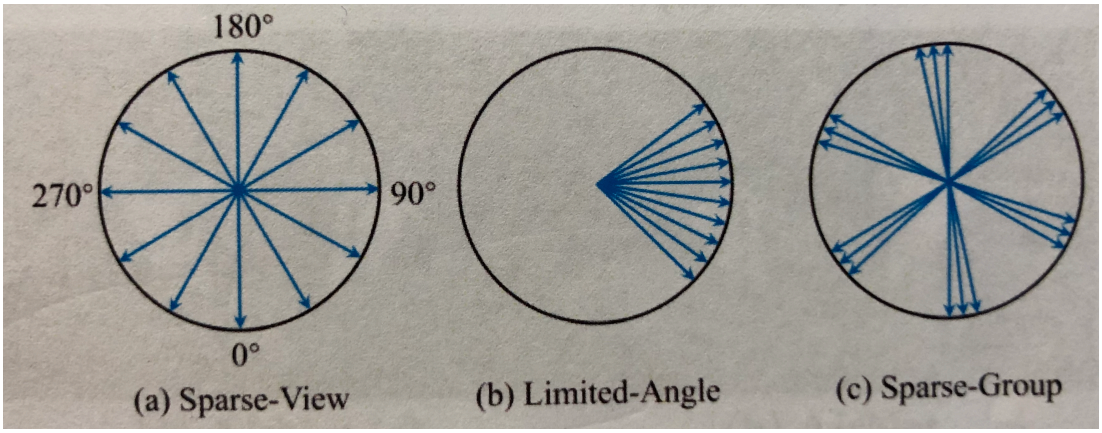


Figure 9 - Geometry of Sparse-View Projections (Chen et al, 2019)

The circles in Figure 9 represent a full 360° rotation around the imaging volume, while the arrows represent specific projection angles. Unlike the sparse-view geometry, the limited-angle geometry only samples a subset of the data but samples at every angle (60°-120° with 1° spacing, for example). In addition, the sparse-group geometry is a collection of limited-angle geometries separated by a larger angular spacing. Ideally, all three geometries would be included in the robustness tests mentioned in the next section, but due to the acquisition geometry (CT-SIM) and optimality of the neural network structure based on previous tests (31), one geometry would sufficient enough to test for robustness.

3.3 Robustness Tests & Evaluation of Robustness Metrics

In general, the robustness of a neural network can be stated as the following problem: Given a neural network with its hyperparameters and training data, is it possible to achieve the same output independent of the input training data or the model hyperparameters?

To apply this problem to AirNET, three neural networks with a given set of hyperparameters were trained based upon the given training libraries (prostate, lung, and abdomen) and were then tested with each of the three libraries. The network parameters were changed six times throughout this process, leading to a total of eighteen neural networks to be analysed. To determine robustness, analyses of the comparison between the predicted CT images and the ground truth CT images were performed with respect to each other. The entirety of training and testing AirNET was performed by utilizing the PyTorch Python library for deep learning.

With respect to image comparison metrics, three different metrics were utilized to evaluate the predicted images from the ground truth images:

3.3.1 Minimum Mean Square Error

Referring back to Equation 15**,** the mean square error calculates the average error of the l_2-norm of the predicted image and the ground truth image. In terms of AirNET, Equation 15 can be rewritten slightly to demonstrate its use in the comparison between the two images:

$$\mathrm{MSE}_{min} = \min\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|I_{pred} - I_{truth}\right\|_2^2\right\} \quad (26)$$

where I accounts for all the pixels in the image in question. Ideally, with noise removed, the mean square error will fall off as the model learns to correctly reconstruct the image data.

3.3.2 Peak Signal-to-Noise Ratio

As one of the most fundamental quantities in image analysis, the signal-to-noise ratio compares the strength of the signal to its overall noise. With respect to x-ray CT, the SNR is defined as follows:

$$\mathrm{SNR} = \frac{\mu}{\sigma} \quad (27)$$

where μ is the average intensity of the image and σ is the standard deviation of the image. In elementary terms, the SNR is simply proportional to the number of x-rays traversing the medium, due to the fact that x-ray counts are Poisson distributed (32). The peak SNR (or PSNR), however, is calculated on a decibel scale (hence the common logarithm in Equation 28, below) due to the fact that most signals and images are composed of a large dynamic range. Therefore, the PSNR is defined to account for such an effect (33):

$$\mathrm{PSNR} = \max\left[20\log\left(\frac{I_{pred}}{\sqrt{\mathrm{MSE}}}\right)\right] \quad (28)$$

It can be concluded that a higher value of the PSNR, the more effective the reconstruction will be. With respect to time, the PSNR increases and then asymptotically approaches its maximum value.

3.3.3 Structural Similarity Index

First described in a 2004 paper (34), the structural similarity index (SSIM) is an image metric that takes into account quantities of luminance, contrast, and optical structure comparison. It was developed in an attempt to define a metric based upon the ability of human visual perception to discern two images apart from one another. Although the derivation of the explicit expression is beyond the scope of this paper, an elementary form of the SSIM is defined below that compares the statistical features of two images *x* and *y*:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (29)$$

where C_1 and C_2 are stability constants that stabilize the weak denominator. As such, the range of values of the SSIM is in the interval [0,1], where 0 represents two non-identical structures and 1 represents two identical structures. For the purposes of AirNET, a value of the SSIM near or extremely close to unity will ensure that to one's perception, two structures in the same image are the same, with a small difference in pixel intensity. Analyses of this metric will be discussed later.

3.4 Image Metric Extraction via TensorFlow

In order to easily extract the relevant image metrics in order to evaluate the efficacy and robustness of AirNET, TensorFlow's API was utilized to conveniently display the time evolution (per epoch) of the image metrics. Plots of each image metric, as well as the step-wise decrease of the learning rate, were shown to monitor the real-time evolution of the model's training. It was also an optimal way to visualize and to determine the maxima of the image metrics in question, with the addition of straightforward data extraction.

CHAPTER 4. RESULTS & DISCUSSIONS

4.1 Image Metric Analysis

As shown in Table 1 below, overall variability of image metrics is dependent on the training library as well as the model hyperparameters utilized. With respect to the training libraries, the abdomen-trained networks have the lowest MSE and have the highest PSNR and SSIM values. Conversely, the lung-trained networks have the highest MSE and the lowest PSNR and SSIM values. This result is most likely due to several different factors both from the image data and the model's hyperparameters.

By looking at a standard CT atlas (35), one can see that in axial slices of the body, the abdomen and pelvis contain major anatomic structures including the pancreas and the kidneys. These organs are mainly composed of tissue and fat, and as a result, appear white on CT images due to large amounts of x-ray attenuation. Viewing the chest, the thoracic anatomy contains many air-filled organs of the respiratory system, which appear black on CT images. If AirNET were to be trained on this data, there would be a larger amount of insufficient data accounted for due to the overall lack of anatomy.

With respect to the model hyperparameters, the results show that as the number of epochs, hidden layers, and iterations increases, the maxima of the image metrics converge and become optimal. This, however, requires a long amount of training time (approximately 2-3 minutes per epoch, total training times range from 3-20 hours). Despite a desire for dense interconnectivity and high amounts of training, a feature that AirNET relies upon, it is possible that AirNET can be over trained, leading to false positive results. An interesting phenomenon occurs with respect to

changing the number of batches of the neural network. Normally, batch normalization, when applied to a neural network, usually improves network speed and reliability. The output, however, contains the smallest values of the PSNR and SSIM and the largest values of the minimum MSE of all the network iterations. This is actually expected, due to the nature of the problem. An expansive classification neural network would benefit from batch normalization, while AirNET may not benefit. A plausible reason for this exists within the training data themselves. Since the training data represents a population of non-unique CT sims of cancer patients, there is a possibility that patients with abnormalities (metal inserts, larger tumors, etc) may be weighted more in the batch normalization. Therefore, the average would be biased towards the image with the greater abnormality. The graphical time dependencies of these image metrics are located in the appendix of this report, along with their convergent maximal or minimal values.

Table 1 - Image Metric Maxima for AirNET (200 Epochs, 48 Layers)

Training Library	Minimum MSE	Peak SNR	SSIM
Prostate	0.05	53.69915	0.99554
Lung	0.103	52.46197	0.992
Abdomen	0.023	57.00737	0.99834

Training Library/Testing Library	Minimum MSE	Peak SNR	SSIM
Prostate/Prostate	0.04975	53.69915	0.99554
Lung/Prostate	0.04641	54.10771	0.9963
Abdomen/Prostate	0.04547	54.38985	0.99652
Prostate/Lung	0.16575	49.91346	0.9883
Lung/Lung	0.10202	52.46197	0.992
Abdomen/Lung	0.11877	52.24656	0.9909
Prostate/Abdomen	0.05316	52.99845	0.99657
Lung/Abdomen	0.0249	56.19546	0.99816
Abdomen/Abdomen	0.02269	57.00738	0.99834

Table 2 - Image Metric Maxima for AirNET (100 Epochs, 24 Layers, 2 Batches)

Training Library	Minimum MSE	Peak SNR	SSIM
Prostate	0.152	49.39493	0.99052
Lung	0.185	48.72467	0.98697
Abdomen	0.099	50.04738	0.9942

Training Library/Testing Library	Minimum MSE	Peak SNR	SSIM
Prostate/Prostate	0.15153	49.39493	0.99052
Lung/Prostate	0.16793	48.49524	0.99026
Abdomen/Prostate	0.20286	47.79984	0.98924
Prostate/Lung	0.40801	45.30737	0.97805
Lung/Lung	0.185	48.72467	0.98697
Abdomen/Lung	0.23632	47.65637	0.98437
Prostate/Abdomen	0.22117	46.92434	0.98984
Lung/Abdomen	0.07628	50.99712	0.99508
Abdomen/Abdomen	0.0991	50.04737	0.9942

4.2 Quantitative Image Differences

To further determine the efficacy of AirNET's reconstruction capabilities, the actual reconstructed CT images were analyzed to detect any minute changes within them. Macroscopic comparisons of the ground truth and reconstructed CT images yield little information into how different the two images are. Taking the absolute difference provides some insight, as shown in the images below. In order to fully display the differences between the ground truth image and the predicted image, manually zooming in to different regions of interest provide a somewhat better view of the differences between both images. In some cases, the normalized maximum intensities of the image pixels differed by nearly 0.02. Though this quantity and others can be calculated by means of computation, visual perception fared better in analyzing the subtle differences between the two images. Samples of such comparisons are presented below.

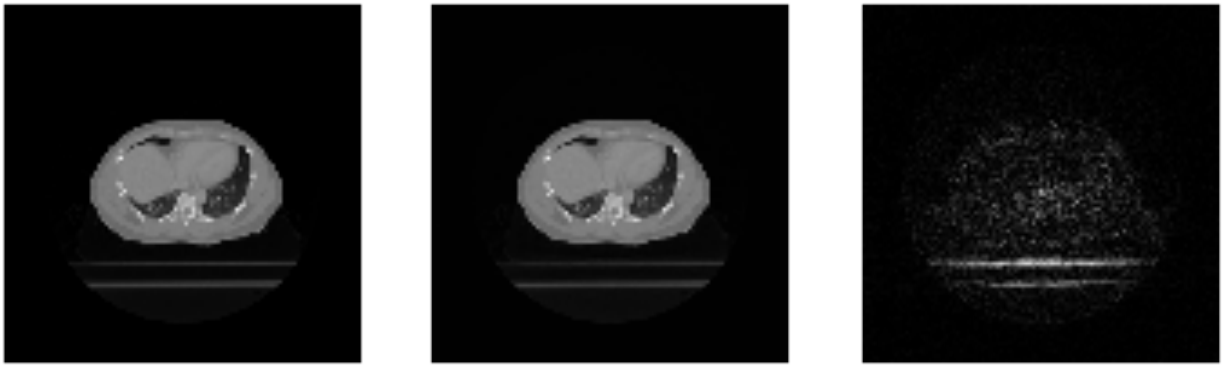


Figure 10 - (L to R) Ground Truth, Predicted, & Difference of Prostate CT Images (100 Epochs, 24 Layers, 2 Batches)

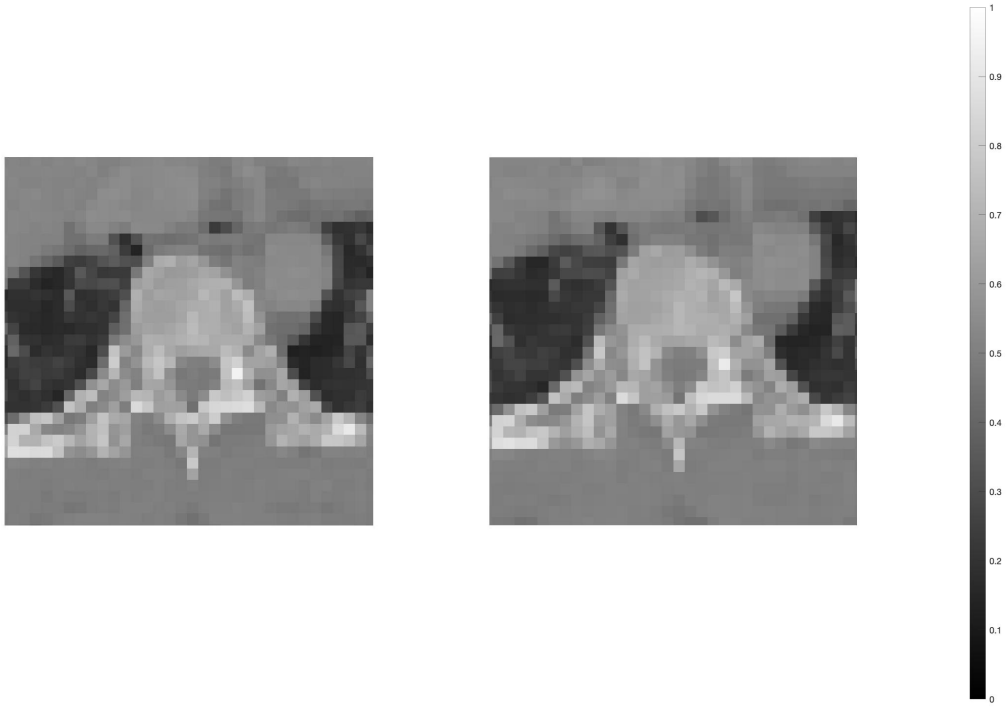


Figure 11 - Zoomed in Axial View of T Spine (Ground Truth Left, Predicted Image Right)

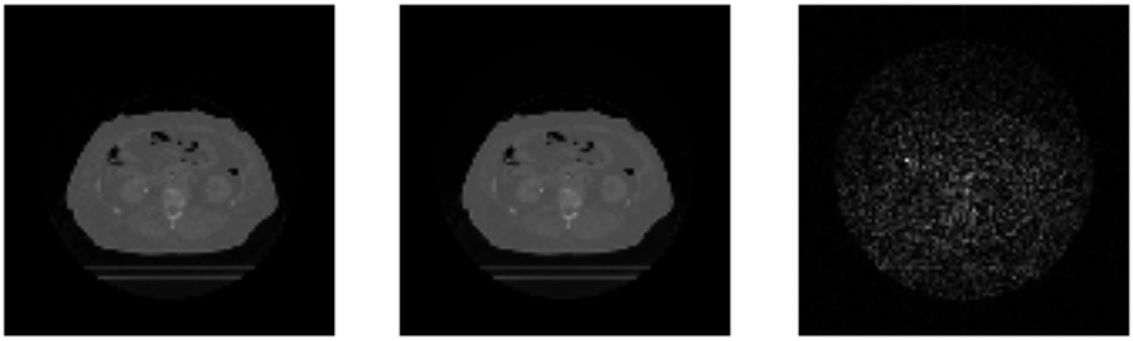


Figure 12 - (L to R) Ground Truth, Predicted, & Difference of Abdominal CT Images (100 Epochs, 24 Layers, 2 Batches)

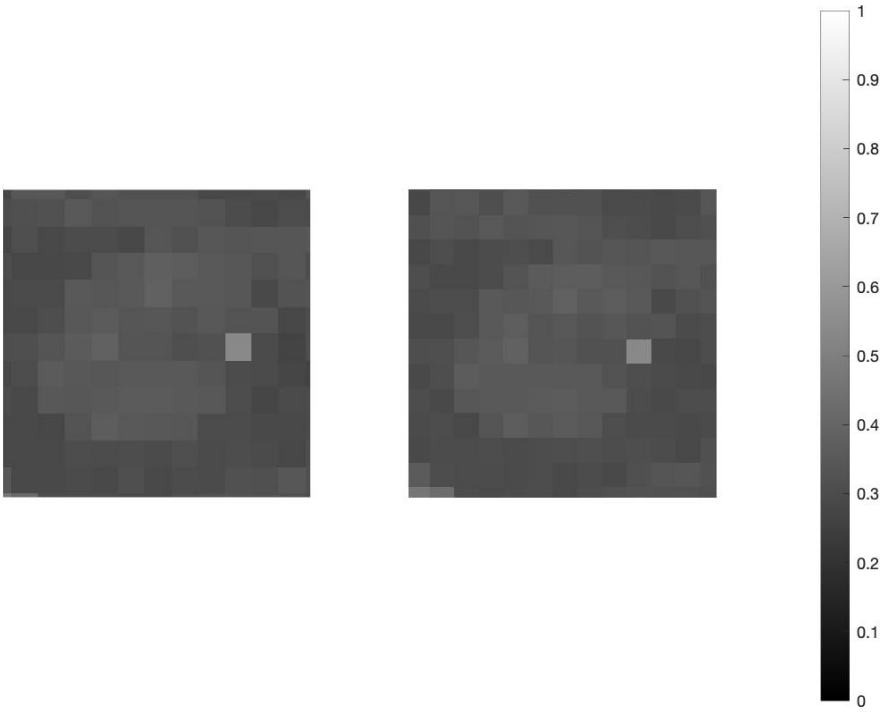


Figure 13 - Zoomed in Axial View of Femoral Head (Ground Truth Left, Predicted Image Right)

Figure 14 - (L to R) Ground Truth, Predicted, & Difference of Lung CT Images (100 Epochs, 24 Layers, 2 Batches)

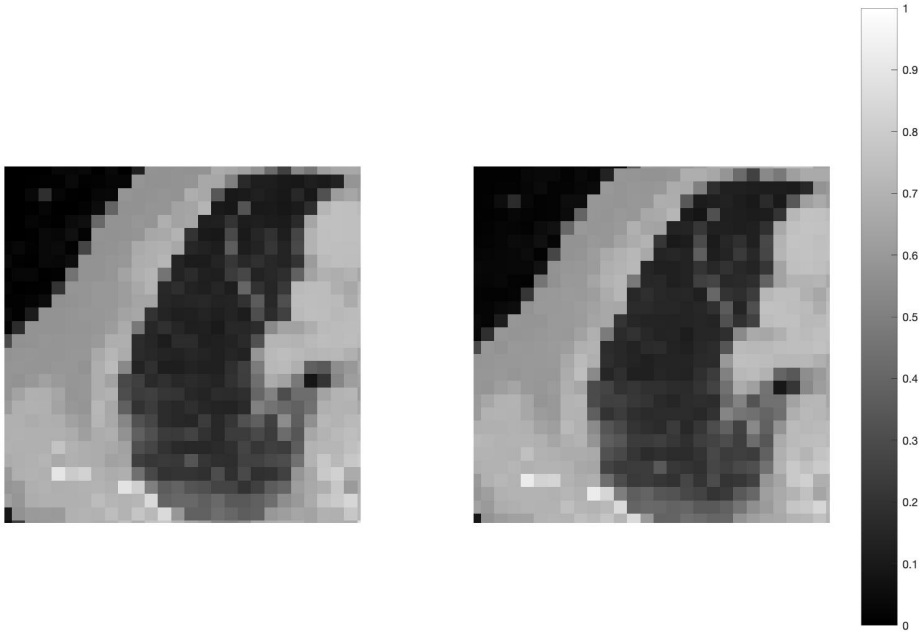


Figure 15 - Zoomed in Axial View of Right Lung (Ground Truth Left, Predicted Image Right)

CHAPTER 5. CONCLUSIONS & FUTURE WORK

From the data obtained over the course of this work, it has become very apparent that AirNET is a very robust neural network for reconstructing sparsely viewed CT images. This is most likely due to two different factors, the first relating to the reconstruction method utilized (AIR), and the second relating to the architecture of AirNET itself. Because AIR involves an "unrolling" type of reconstruction method, it combines the positive aspects of filtered-back projection and iterative techniques to ensure sufficient reconstruction results. Furthermore, AIR is fairly computationally efficient and does not require a significant amount of computational time to process.

Even more so than the reconstruction method involved in the reconstruction, the structure of AirNET itself justifies why such a method is very useful in both the reconstruction of images and the evaluation of the model's robustness. The convolutional nature of the network, along with the fact that AirNET is densely connected, ensures that all image features are extracted during the reconstruction process and are then applied to the model with each passing iteration. The application of several convolutional-ReLU layers further help the process of reconstruction, and the resulting images that are predicted look nearly like the ground truth images.

With respect to the training library, it has been shown that AirNET is very robust with respect to the training libraries used and the appropriate model parameters. On a broad scale, these results are sufficient, however more robustness tests beyond the scope of this report would be required to examine all plausible training libraries. For example, a library of head/neck cancer patients would possibly have a more difficult time than predicting an image of the pelvic area. A second and more extreme example would be utilizing a library that contains a CT artifact, such as beam hardening due to a metal insert inside a patient to a ring artifact caused by a broken detector. It is possible

that such effects may disrupt the image reconstruction, leading to many false positive predictions. There is some work (40) that attempts to remove any possible artifacts and effects of noise, but more improvements are currently being studied.

To extend this work further, there are a few possible pathways to take. The first pathway concerns changing the neural network architecture in such a way so that more information is able to be extracted while computational cost is kept constant. Implementing AIR as a reconstructive means into the architecture of several networks in the literature (36-42) would be of interest, both in terms of computational efficiency and in terms of resulting image reconstruction. The second pathway is to apply such neural networks to other imaging modalities, such as MR and ultrasound. Similar network architectures can be found (41) (42) that attempt to reconstruct such networks, but the reconstruction techniques required are unique to the imaging modality itself. Nevertheless, neural networks will become a mainstay in sparse image reconstruction and will be utilized to their fullest extent.

APPENDIX

A) Image Metrics for Robustness Tests

The following tables summarize the minimum mean square error, Peak SNR, and SSIM values for each of the remaining cases.

400 Epochs, 48 Layers:

Training Library	Minimum MSE	Peak SNR	SSIM
Prostate	0.2696	56.93	0.9974
Lung	0.03995	52.63	0.9921
Abdomen	0.2025	57.59	0.9985

Training Library/Testing Library	Minimum MSE	Peak SNR	SSIM
Prostate/Prostate	0.03108	56.11189	0.99738
Lung/Prostate	0.04708	54.07344	0.99643
Abdomen/Prostate	0.04053	54.92963	0.99685
Prostate/Lung	0.42194	50.50762	0.98728
Lung/Lung	0.10512	52.53892	0.99196
Abdomen/Lung	0.11417	52.5698	0.99216
Prostate/Abdomen	0.04261	55.15072	0.99764
Lung/Abdomen	0.02281	56.71811	0.99832
Abdomen/Abdomen	0.02082	57.59508	0.99849

200 Epochs, 24 Layers:

Training Library	Minimum MSE	Peak SNR	SSIM
Prostate	0.039	55.00249	0.99672
Lung	0.109	52.22465	0.99173
Abdomen	0.028	56.03278	0.99803

Training Library/Testing Library	Minimum MSE	Peak SNR	SSIM
Prostate/Prostate	0.03852	55.0025	0.99672
Lung/Prostate	0.04762	53.92403	0.99611
Abdomen/Prostate	0.05456	53.60669	0.99599
Prostate/Lung	0.1692	50.29671	0.98866
Lung/Lung	0.04762	52.22465	0.99611
Abdomen/Lung	0.13326	51.68743	0.99014
Prostate/Abdomen	0.04571	54.05165	0.99732
Lung/Abdomen	0.02628	55.90878	0.99806
Abdomen/Abdomen	0.02763	56.03276	0.99803

100 Epochs, 24 Layers:

Training Library	Minimum MSE	Peak SNR	Peak SSIM
Prostate	0.071	52.5134	0.9948
Lung	0.125	51.11219	0.99472
Abdomen	0.043	53.91585	0.99716

Training Library/Testing Library	Minimum MSE	Peak SNR	Peak SSIM
Prostate/Prostate	0.7115	52.5134	0.9948
Lung/Prostate	0.0704	52.17007	0.99472
Abdomen/Prostate	0.08238	51.66843	0.99439
Prostate/Lung	0.22981	48.28207	0.98577
Lung/Lung	0.12492	55.11219	0.9905
Abdomen/Lung	0.14843	50.38773	0.98903
Prostate/Abdomen	0.8868	50.93157	0.9953
Lung/Abdomen	0.03754	54.22952	0.99733
Abdomen/Abdomen	0.04254	53.91585	0.99716

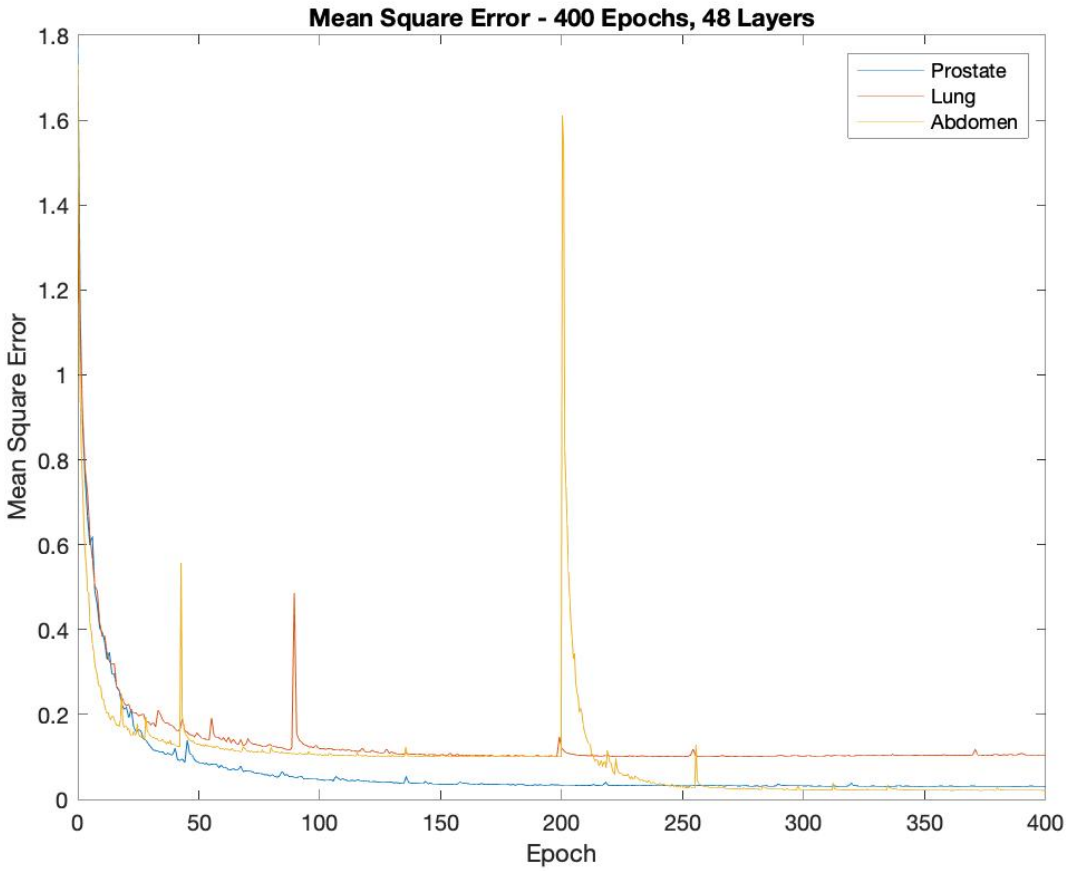
100 Epochs, 24 Layers, 100 Iterations:

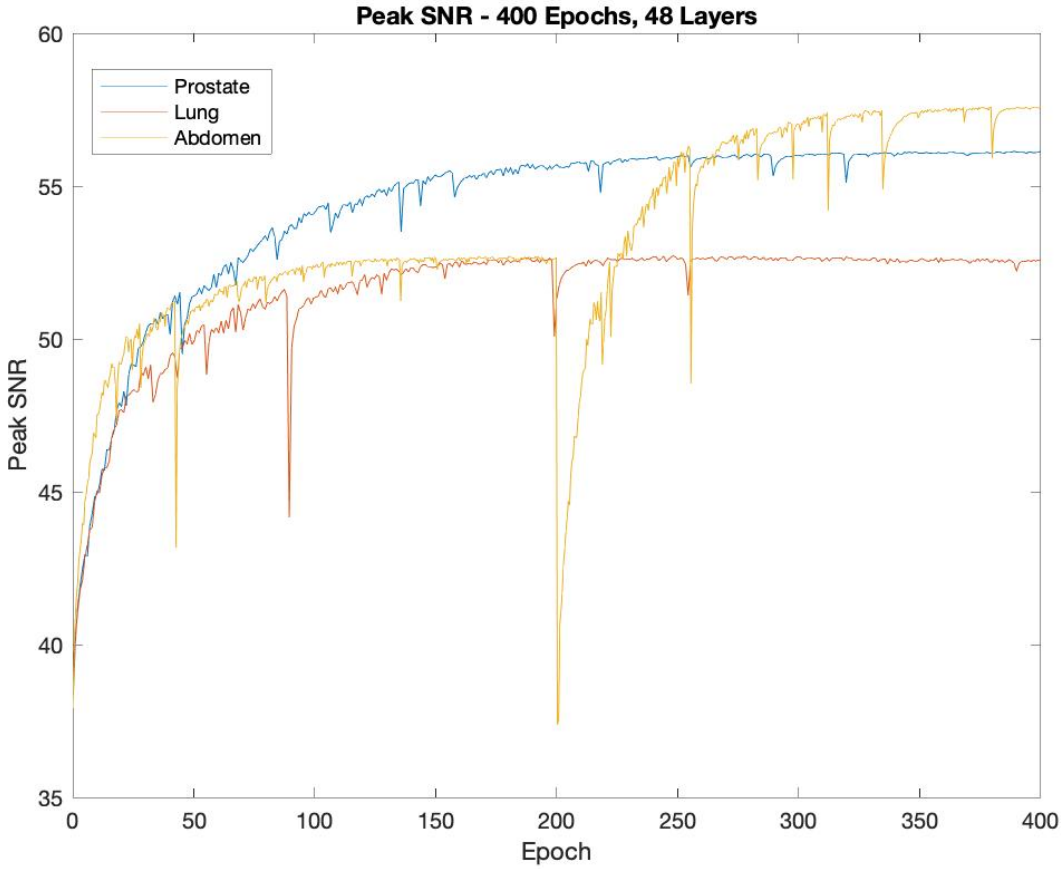
Training Library	Minimum MSE	Peak SNR	Peak SSIM
Prostate	0.044	54.58093	0.99653
Lung	0.109	51.99953	0.99154
Abdomen	0.027	56.26599	0.9981

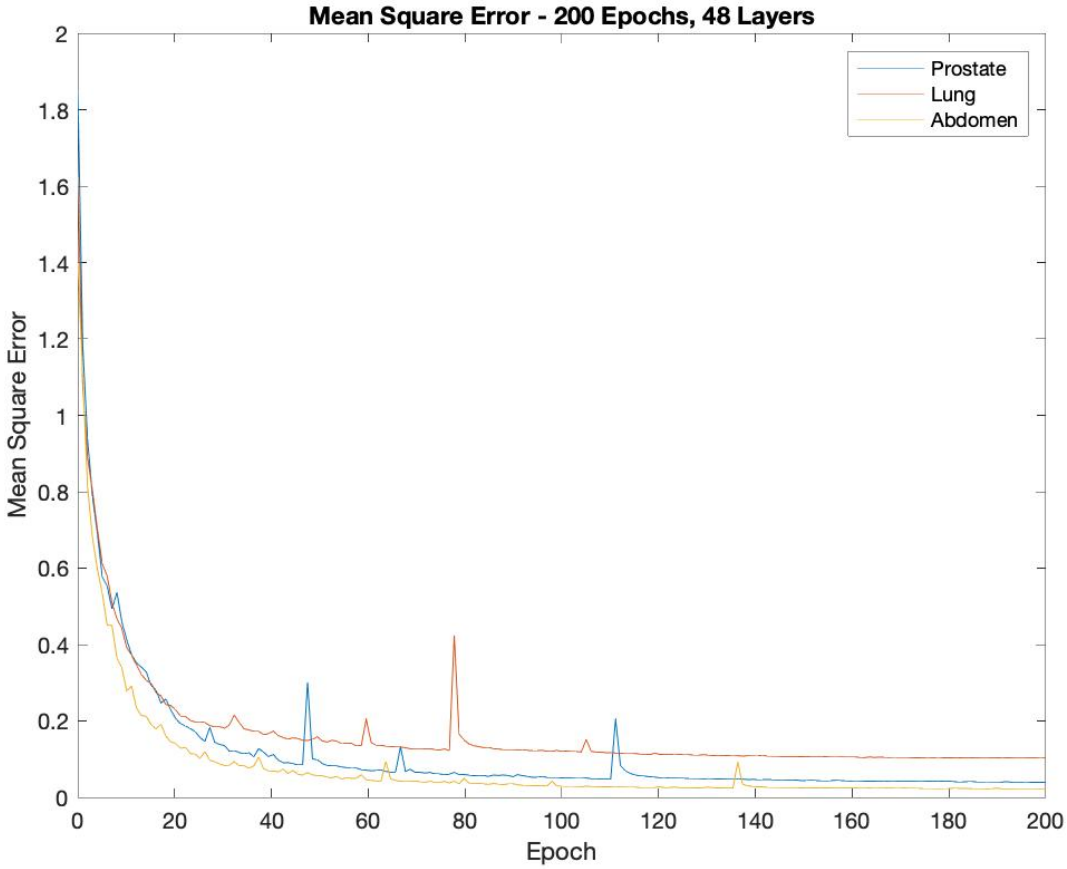
Training Library/Testing Library	Minimum MSE	Peak SNR	Peak SSIM
Prostate/Prostate	0.04383	54.58093	0.99653
Lung/Prostate	0.05373	53.40643	0.9958
Abdomen/Prostate	0.05178	53.83057	0.99614
Prostate/Lung	0.18235	49.76924	0.98844
Lung/Lung	0.10872	51.99953	0.99154
Abdomen/Lung	0.12553	51.8389	0.99056
Prostate/Abdomen	0.05652	53.27611	0.99698
Lung/Abdomen	0.02902	55.4436	0.99783
Abdomen/Abdomen	0.02727	56.26599	0.9981

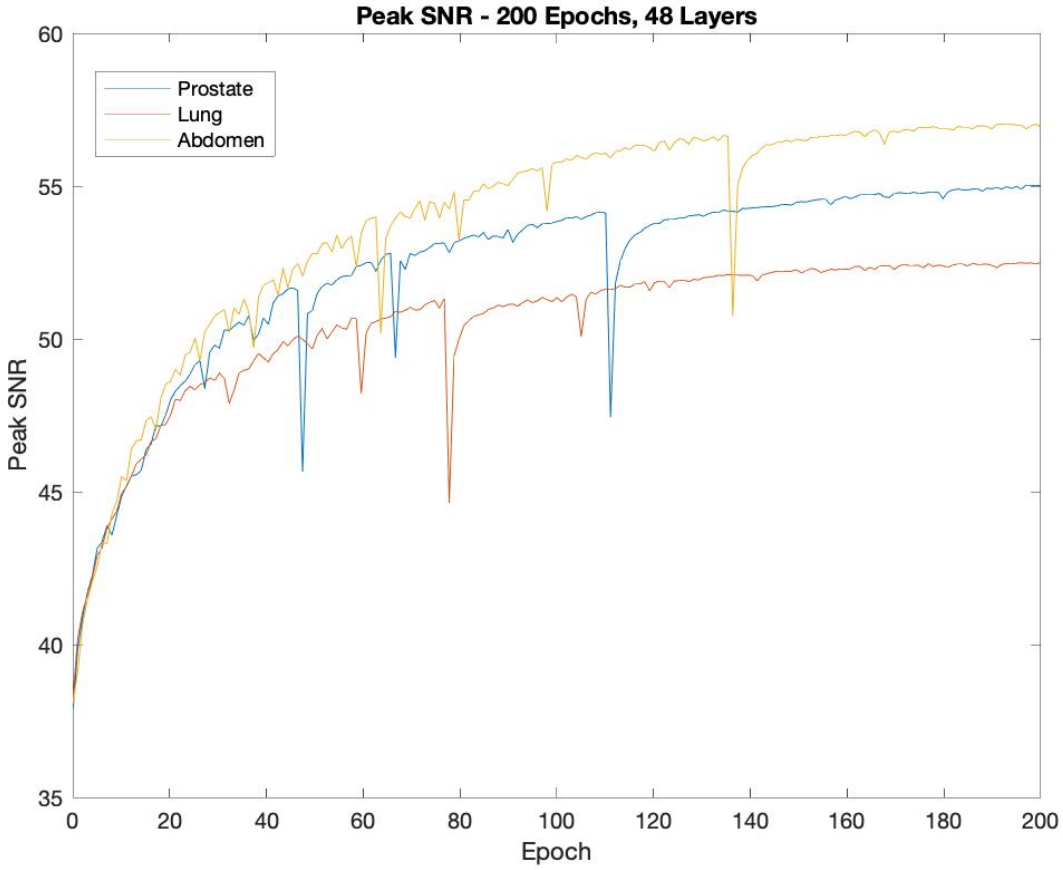
B) Time Dependencies of Image Metrics

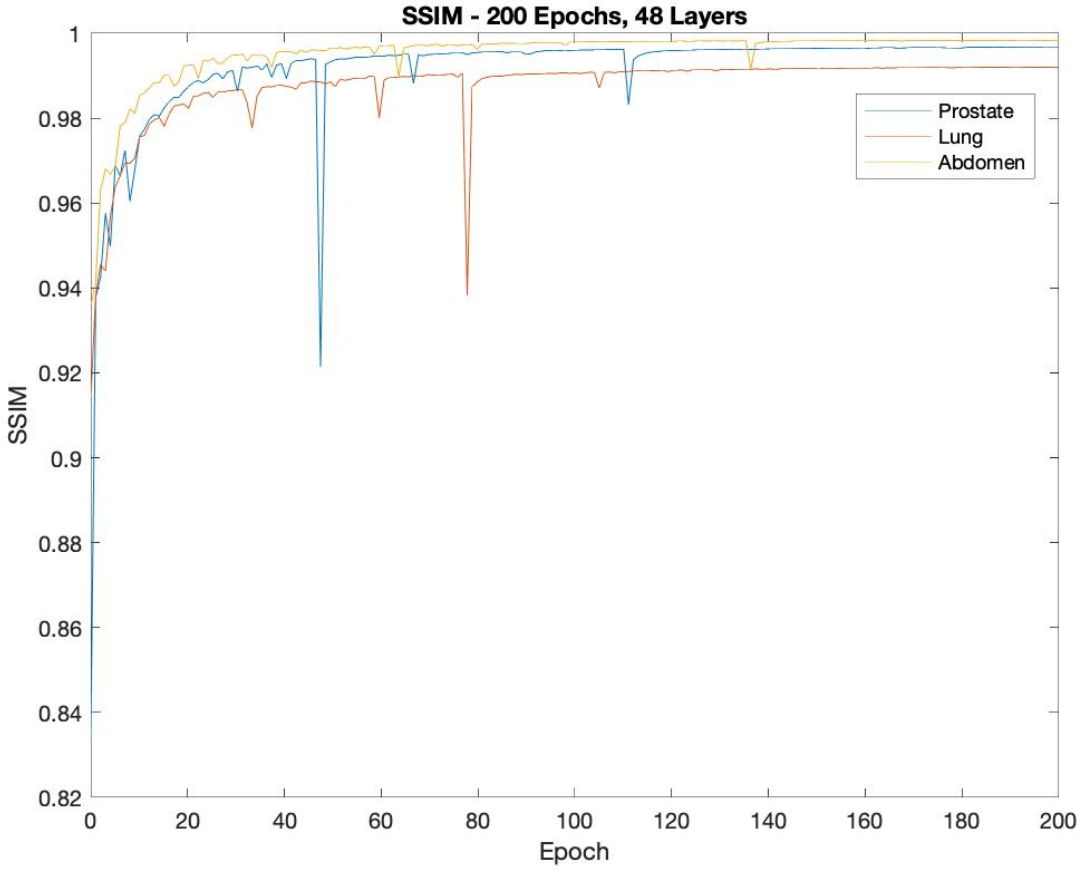
The following eighteen plots depict the time dependence (per epoch) of the minimum mean square error, peak SNR, and SSIM for each of the network parameters above. The specific networks are described in the title of each plot.

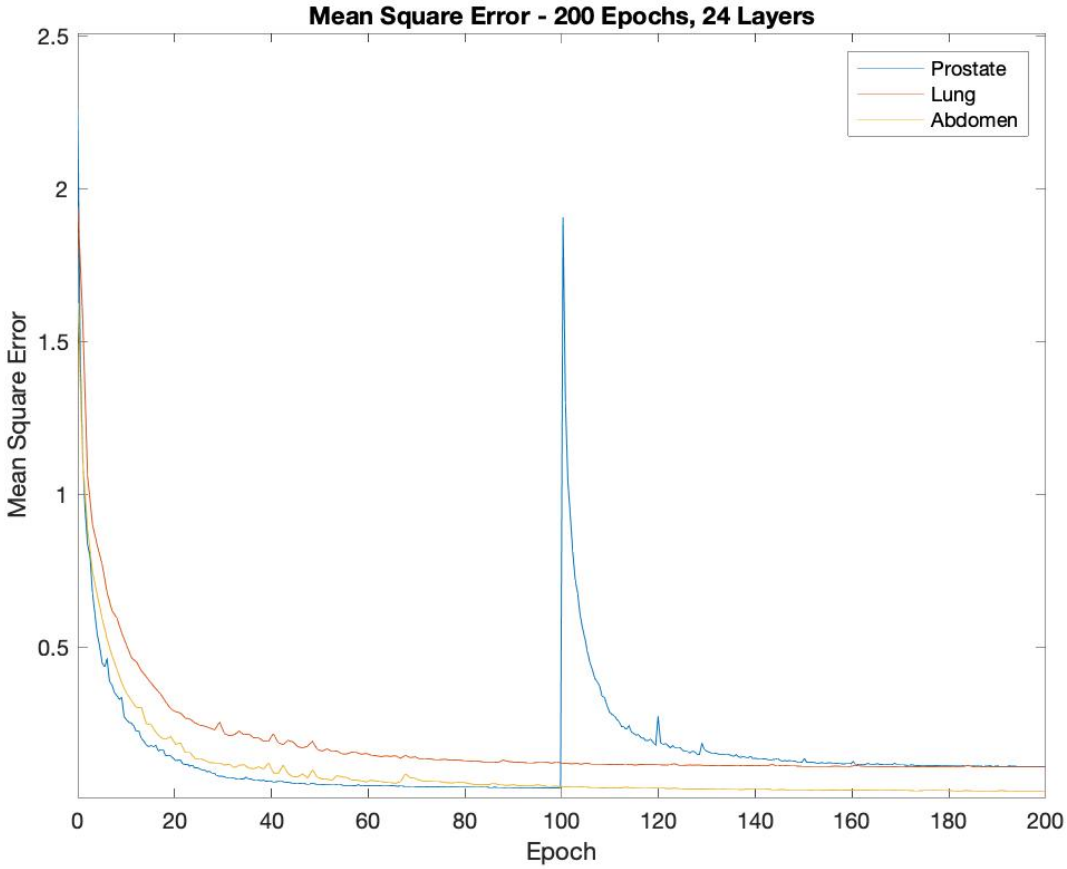


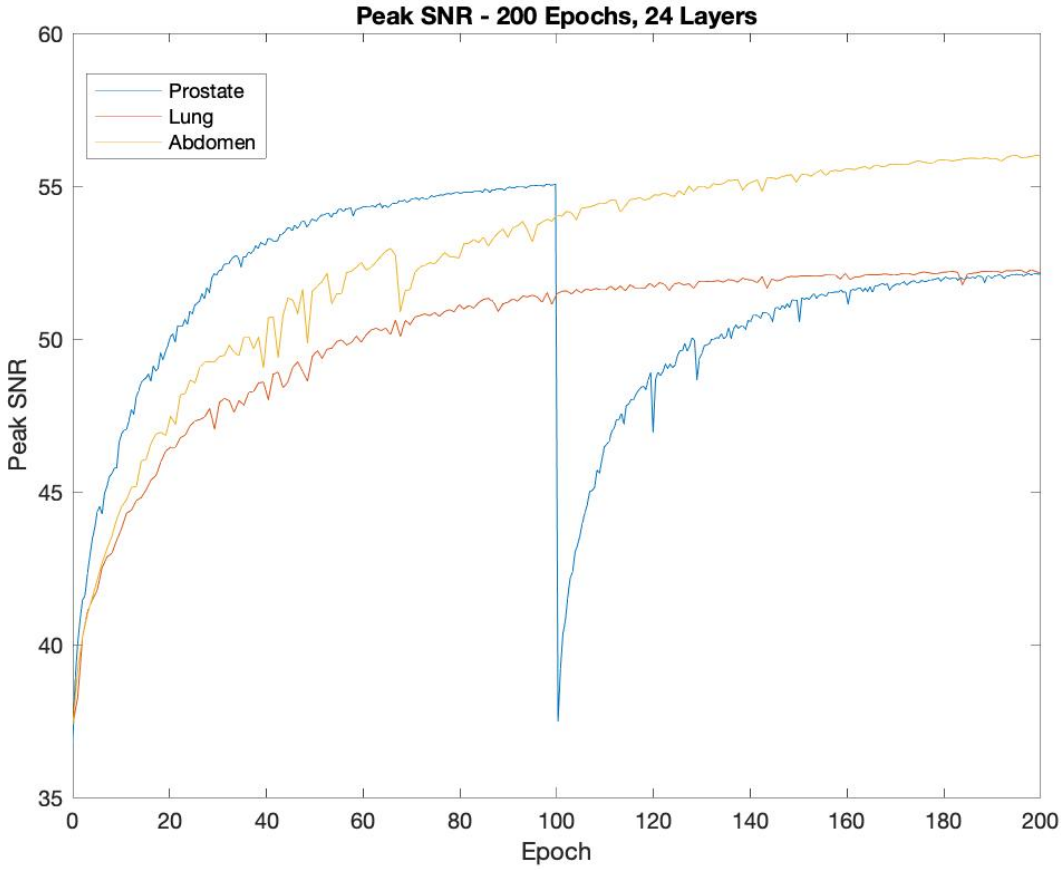

Peak SNR - 400 Epochs, 48 Layers
Prostate
Lung
Abdomen
Peak SNR
Epoch
60
55
50
45
40
35
0
50
100
150
200
250
300
350
400

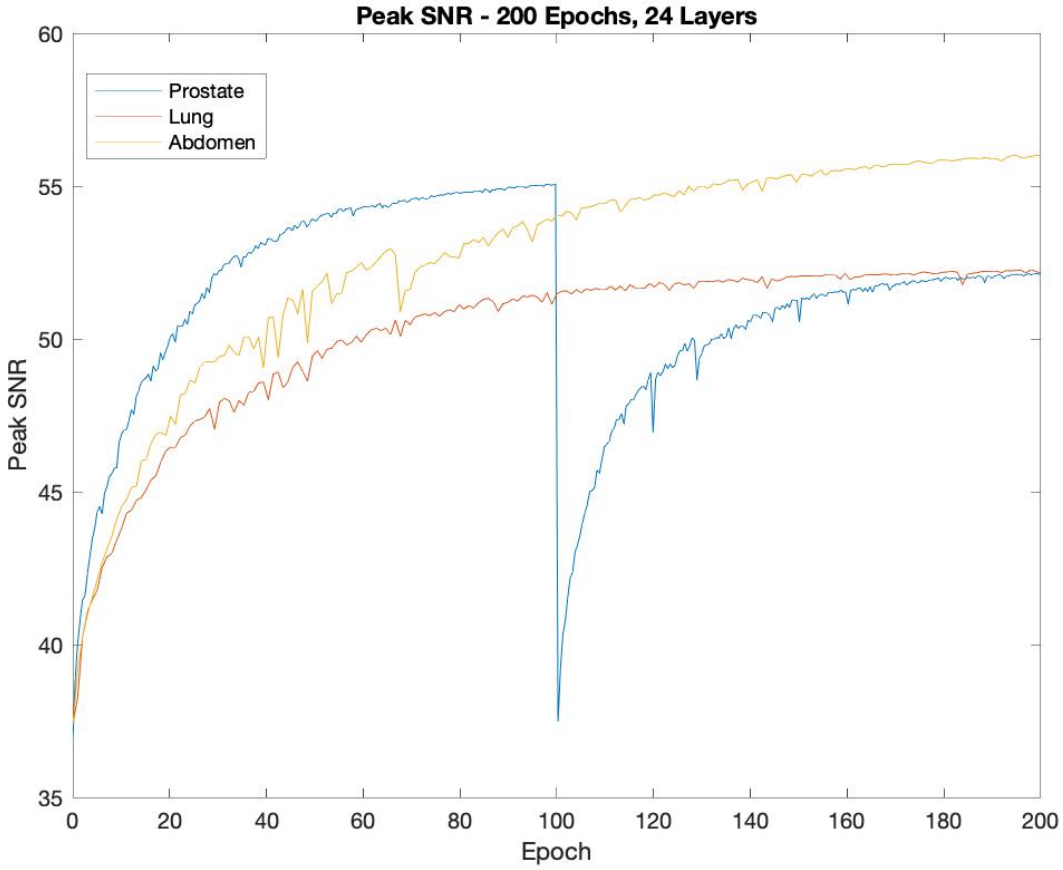

Mean Square Error - 200 Epochs, 48 Layers
Prostate
Lung
Abdomen
Mean Square Error
Epoch
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120
140
160
180
200

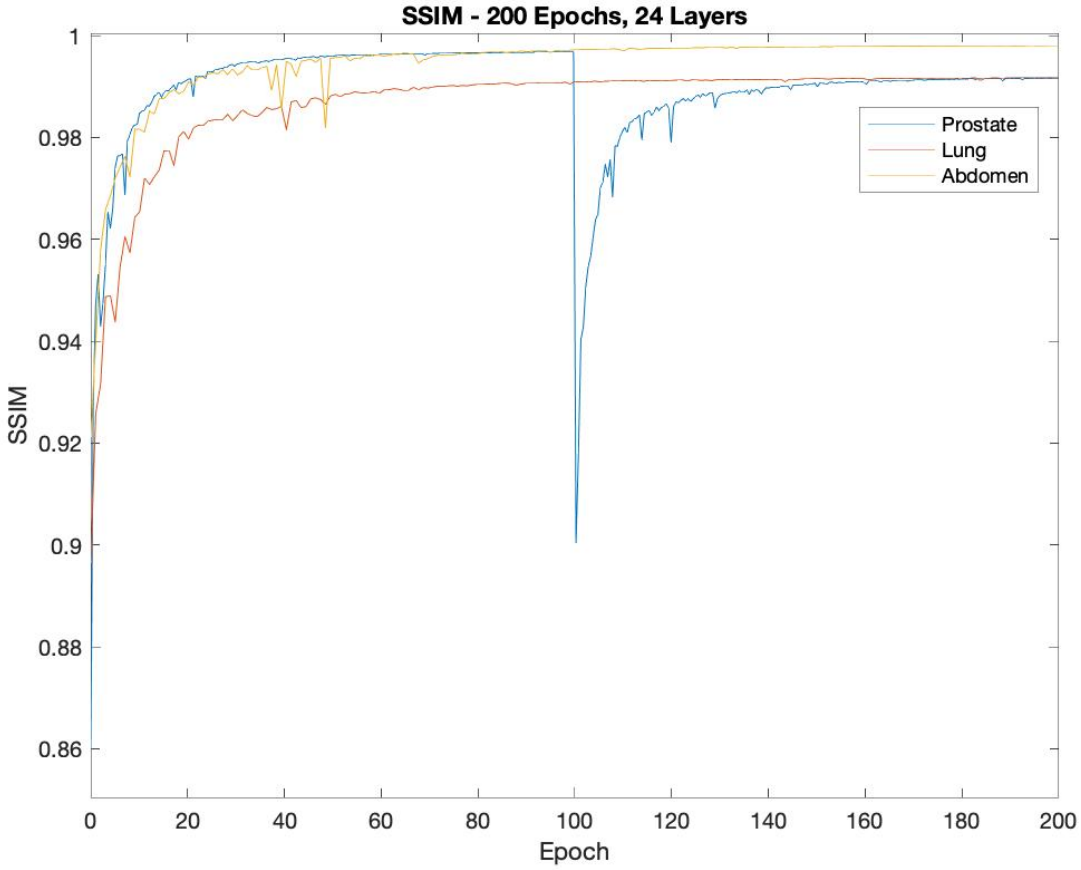

Peak SNR - 200 Epochs, 48 Layers
Prostate
Lung
Abdomen
Peak SNR
Epoch
60
55
50
45
40
35
0
20
40
60
80
100
120
140
160
180
200

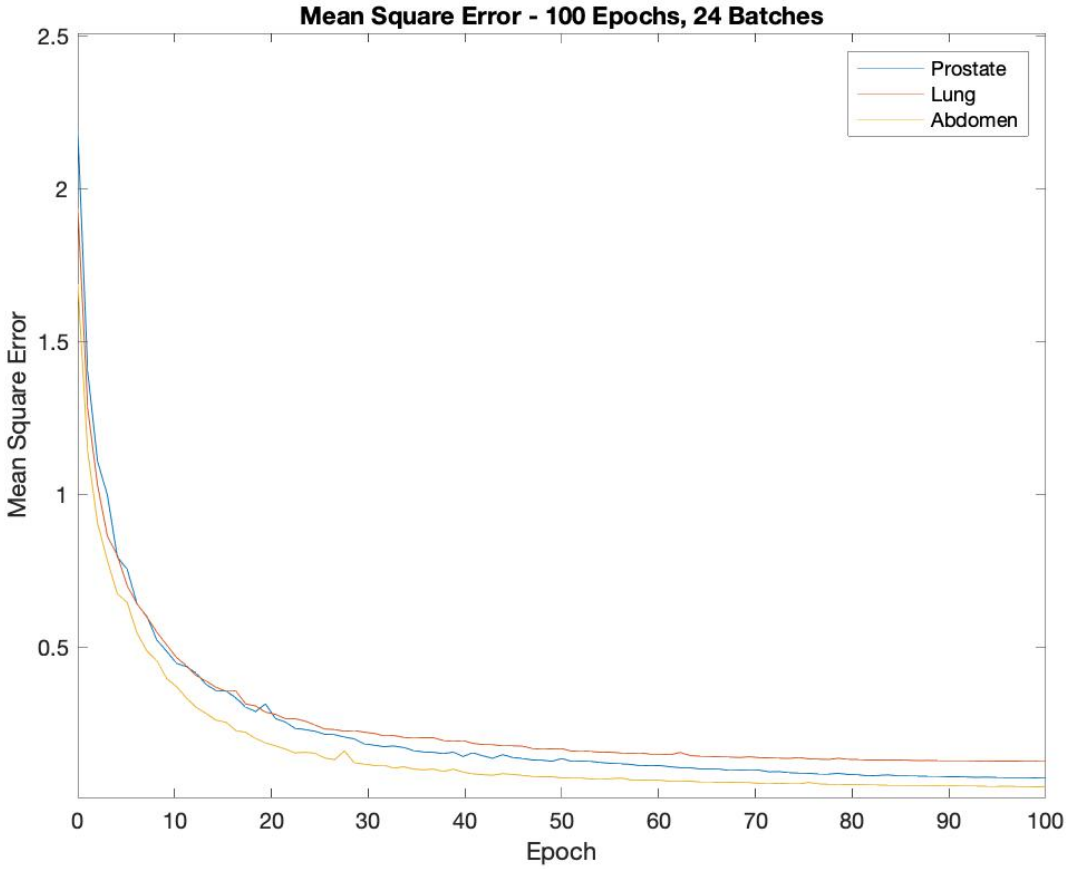

SSIM - 200 Epochs, 48 Layers
SSIM
Epoch
Prostate
Lung
Abdomen
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0
20
40
60
80
100
120
140
160
180
200

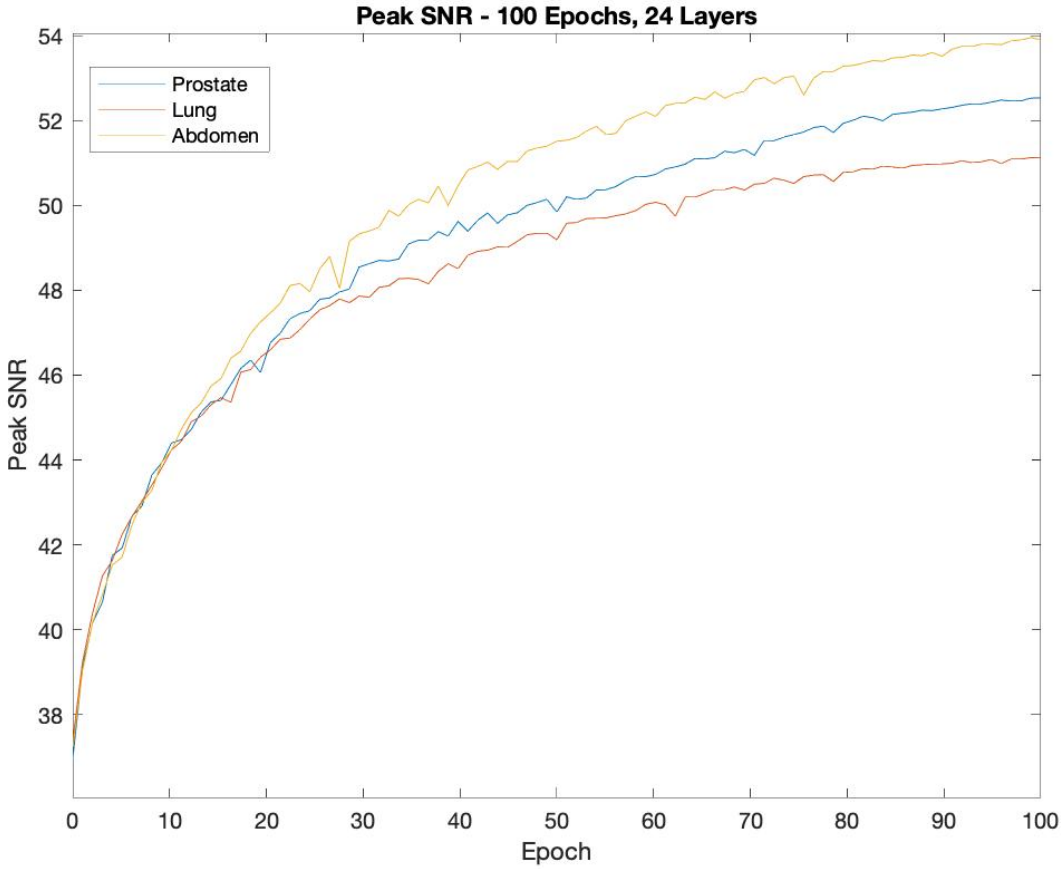

Mean Square Error - 200 Epochs, 24 Layers
Prostate
Lung
Abdomen
Mean Square Error
Epoch
2.5
2
1.5
1
0.5
0
20
40
60
80
100
120
140
160
180
200

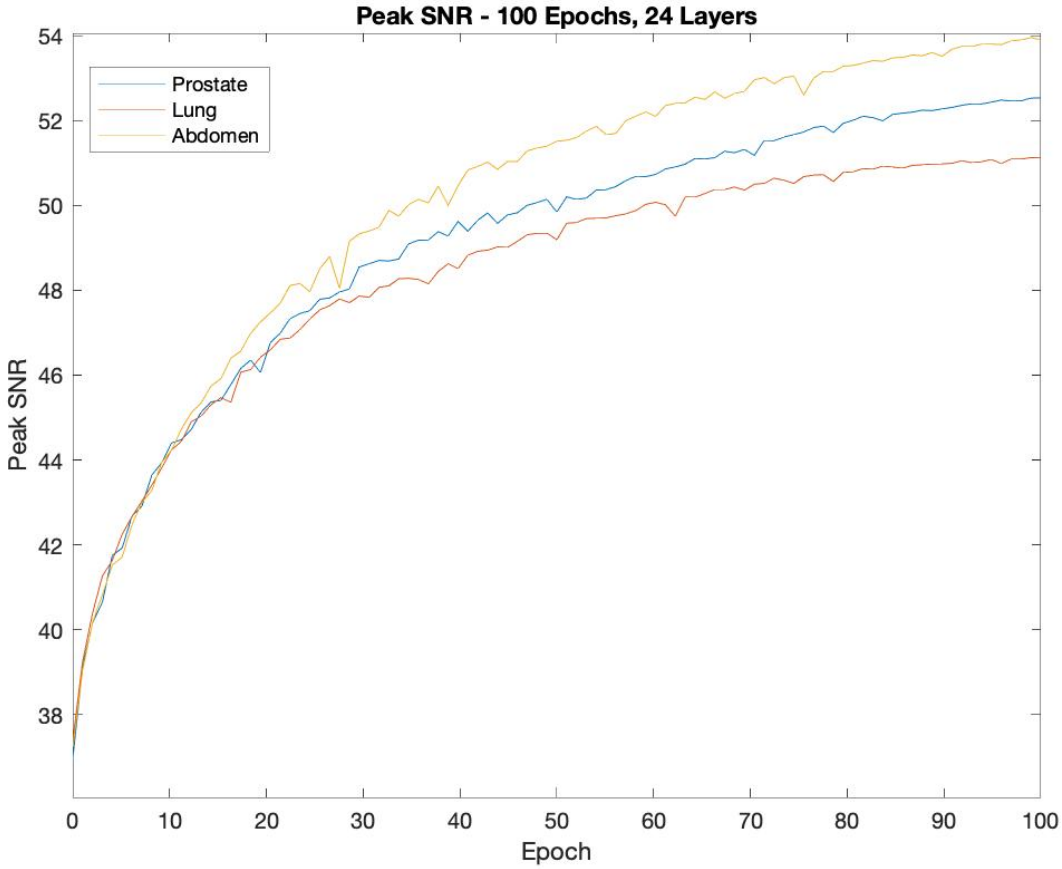

Peak SNR - 200 Epochs, 24 Layers
Prostate
Lung
Abdomen
Peak SNR
Epoch
60
55
50
45
40
35
0
20
40
60
80
100
120
140
160
180
200

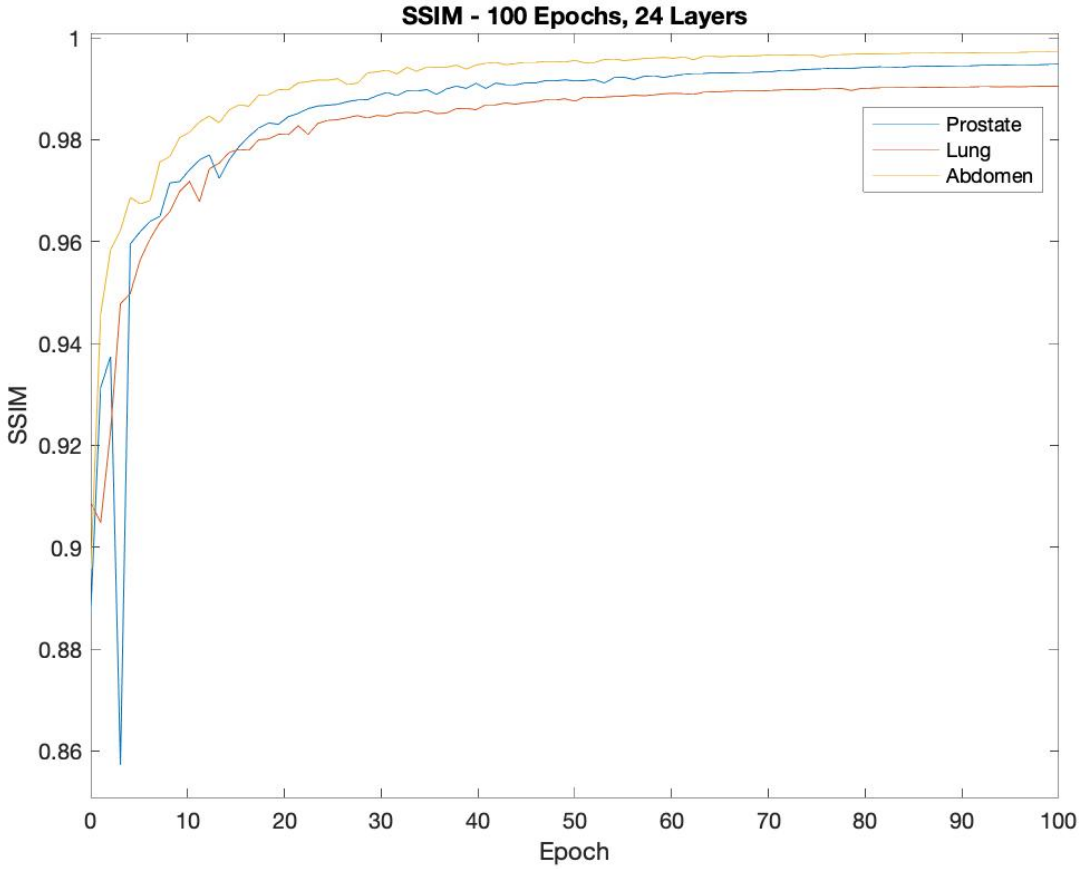

Peak SNR - 200 Epochs, 24 Layers
Prostate
Lung
Abdomen
Peak SNR
Epoch
60
55
50
45
40
35
0
20
40
60
80
100
120
140
160
180
200

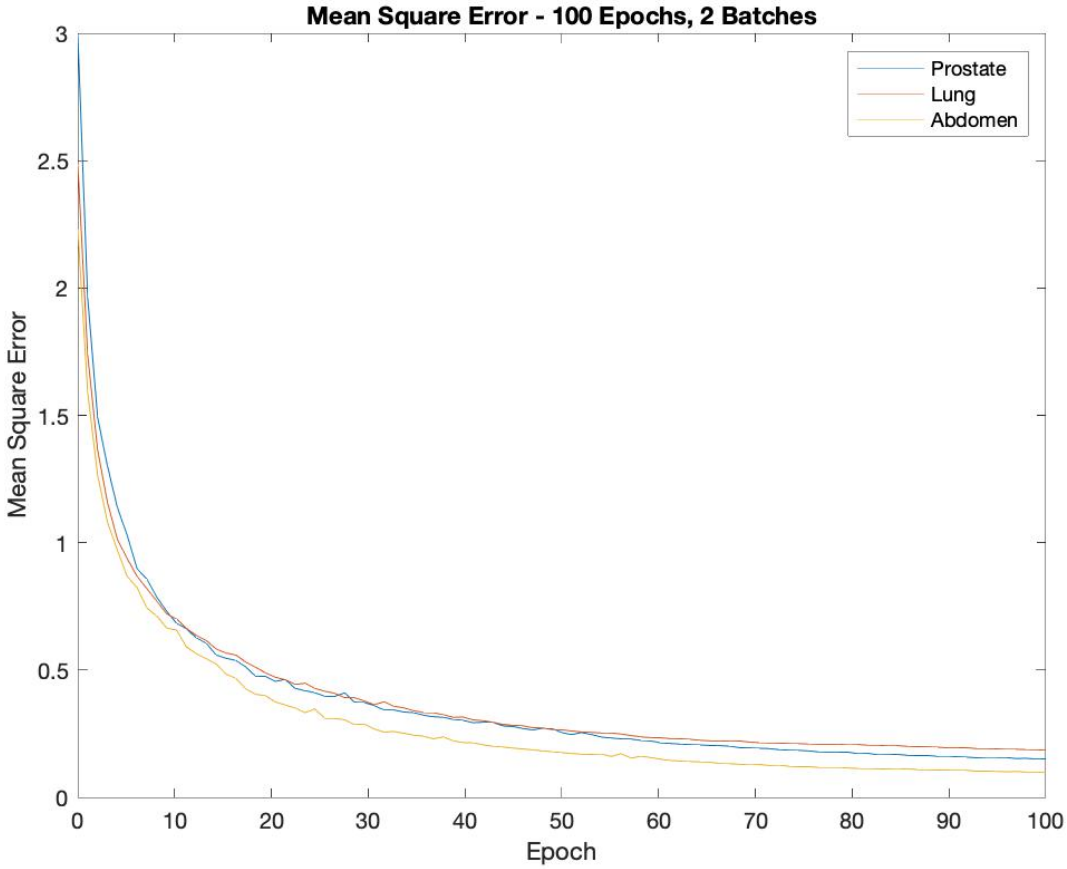

SSIM - 200 Epochs, 24 Layers
SSIM
Epoch
Prostate
Lung
Abdomen
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0
20
40
60
80
100
120
140
160
180
200

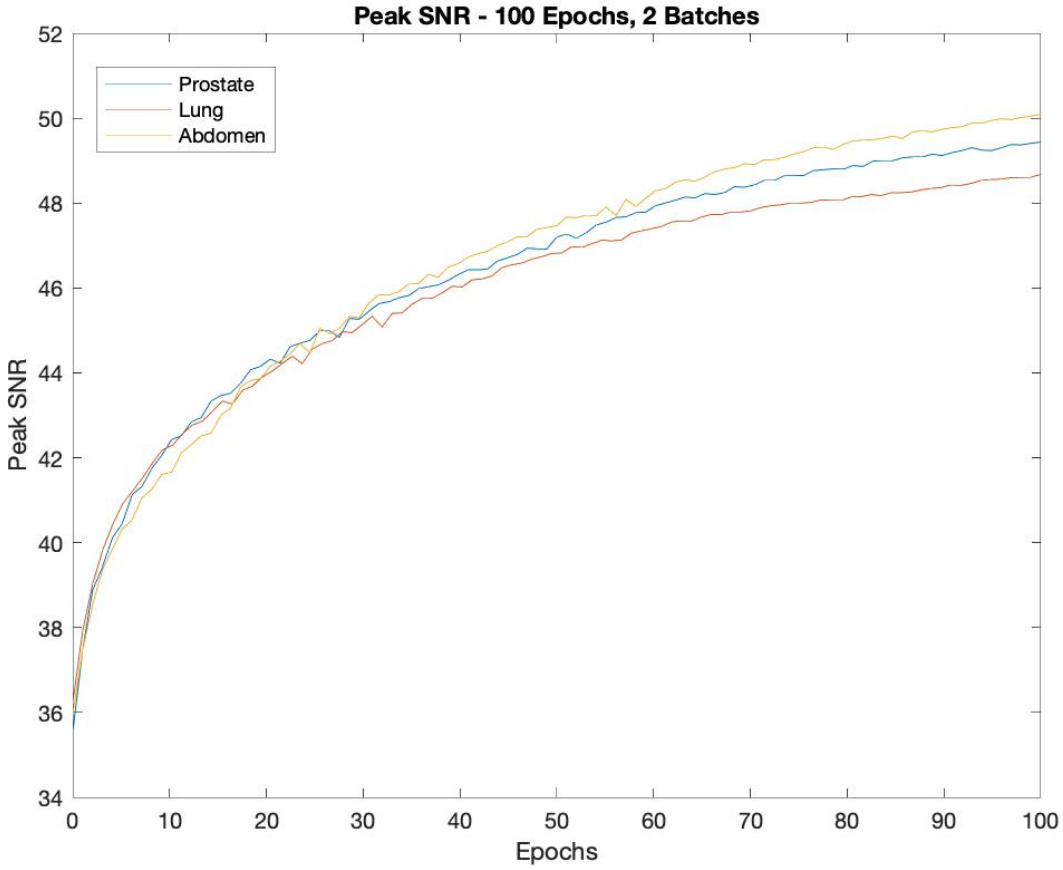

Mean Square Error - 100 Epochs, 24 Batches
Prostate
Lung
Abdomen
Mean Square Error
Epoch
2.5
2
1.5
1
0.5
0
10
20
30
40
50
60
70
80
90
100

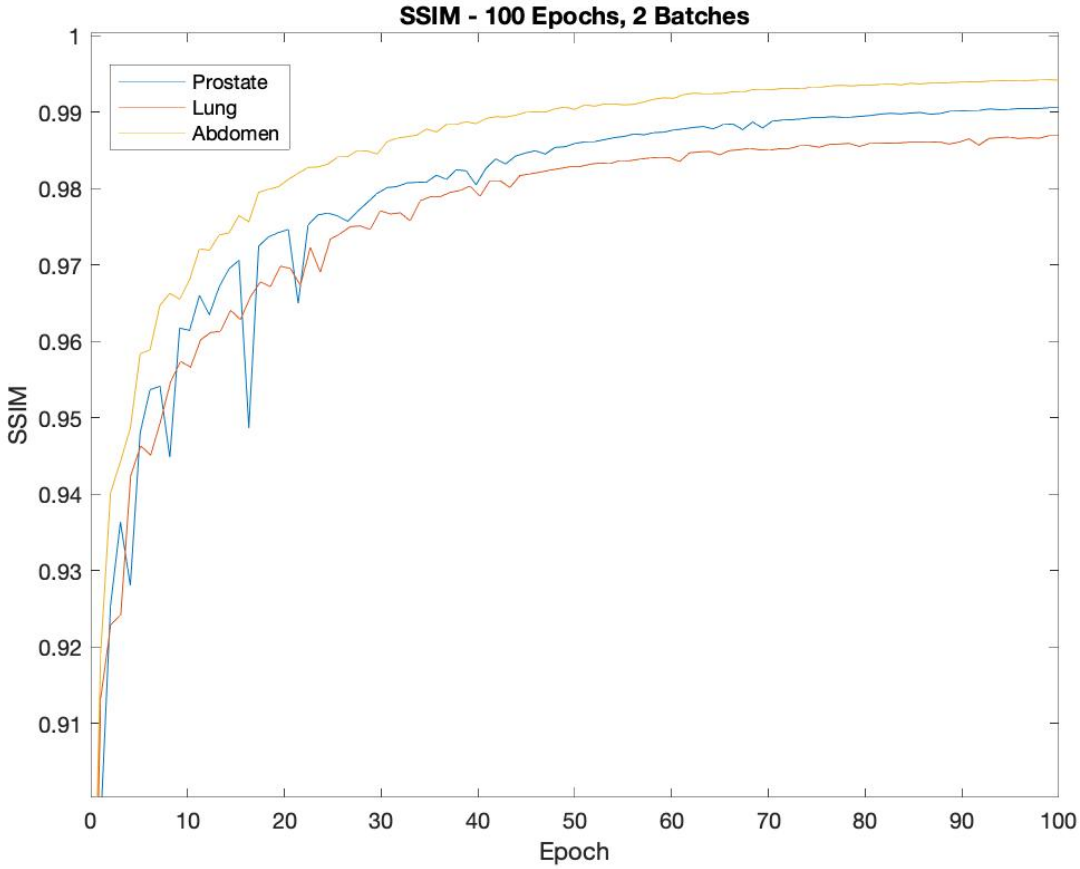

Peak SNR - 100 Epochs, 24 Layers
Prostate
Lung
Abdomen
Peak SNR
Epoch
54
52
50
48
46
44
42
40
38
0
10
20
30
40
50
60
70
80
90
100

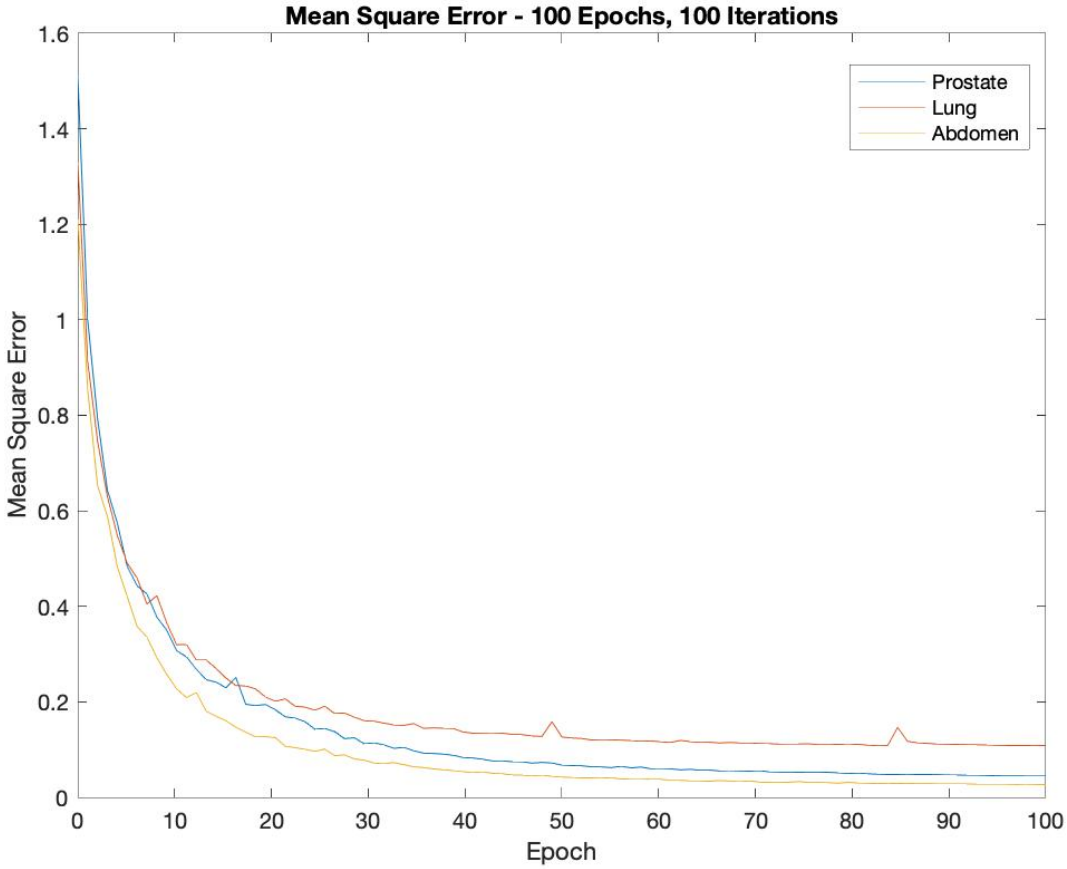

Peak SNR - 100 Epochs, 24 Layers
Prostate
Lung
Abdomen
Peak SNR
Epoch
54
52
50
48
46
44
42
40
38
0
10
20
30
40
50
60
70
80
90
100

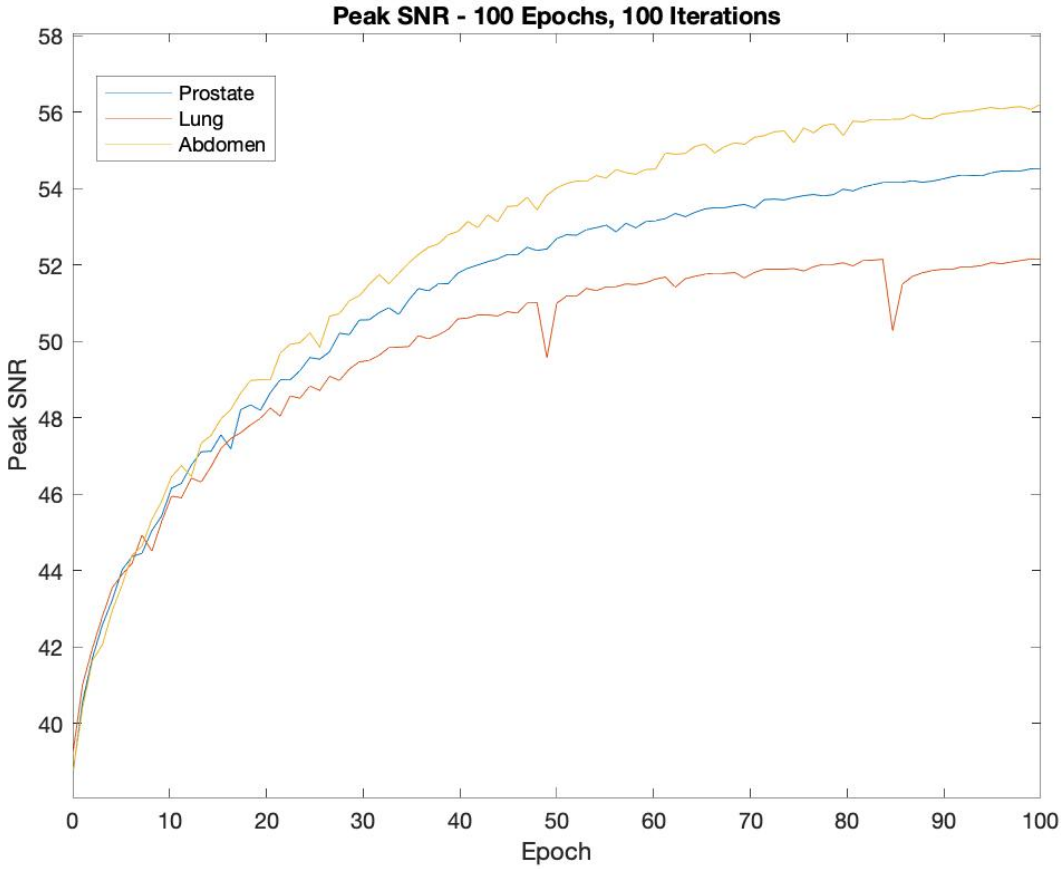

SSIM - 100 Epochs, 24 Layers
SSIM
Epoch
Prostate
Lung
Abdomen
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0
10
20
30
40
50
60
70
80
90
100

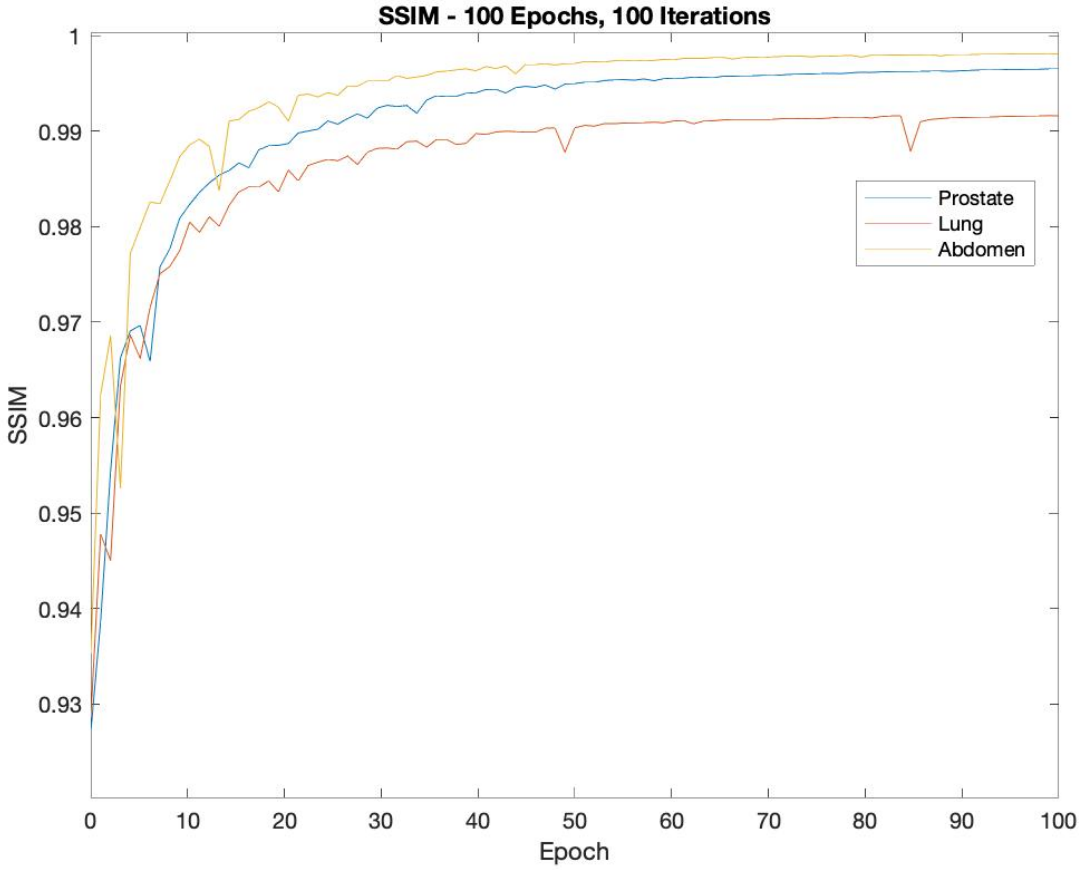

Mean Square Error - 100 Epochs, 2 Batches
Prostate
Lung
Abdomen
Mean Square Error
Epoch
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60
70
80
90
100


Peak SNR - 100 Epochs, 2 Batches
Prostate
Lung
Abdomen
Peak SNR
Epochs
52
50
48
46
44
42
40
38
36
34
0
10
20
30
40
50
60
70
80
90
100


SSIM - 100 Epochs, 2 Batches
Prostate
Lung
Abdomen
SSIM
Epoch
1
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0
10
20
30
40
50
60
70
80
90
100


Mean Square Error - 100 Epochs, 100 Iterations
Prostate
Lung
Abdomen
Mean Square Error
Epoch
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
10
20
30
40
50
60
70
80
90
100


Peak SNR - 100 Epochs, 100 Iterations
Prostate
Lung
Abdomen
Peak SNR
Epoch
58
56
54
52
50
48
46
44
42
40
0
10
20
30
40
50
60
70
80
90
100


SSIM - 100 Epochs, 100 Iterations
SSIM
Epoch
Prostate
Lung
Abdomen
1
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0
10
20
30
40
50
60
70
80
90
100

REFERENCES

[1] Smith-Bindman R, Kwan ML, Marlow EC, et al. "Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada", 2000-2016. *JAMA.* 2019; 322(9):843–856.

[2] National Council on Radiation Protection and Measurements. *Report No. 160, Ionizing Radiation Exposure of the Population of the United States.* Bethesda, MD: NCRP, 2009.

[3] Agasthya, Greeshma A., et al. "Can Breast Compression Be Reduced in Digital Mammography and Breast Tomosynthesis?" *American Journal of Roentgenology*, vol. 209, no. 5, Nov. 2017.

[4] Candès, Emmanuel J., et al. "Stable Signal Recovery from Incomplete and Inaccurate Measurements". *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, 2006, pp. 1207–1223.

[5] Donoho, D.l. "Compressed Sensing". IEEE Transactions on Information Theory, vol. 52, no. 4, 2006, pp. 1289–1306.

[6] Domingues, Inês, et al. "Using Deep Learning Techniques in Medical Imaging: A Systematic Review of Applications on CT and PET". *Artificial Intelligence Review*, 2019.

[7] Gonzalez, A. Berrington De, et al. "Projected Cancer Risks from Computed Tomographic Scans Performed in the United States in 2007". *Journal of Vascular Surgery,* vol. 51, no. 3, 2010, p. 783.

[8] Evans, Robley D. *The Atomic Nucleus*. McGraw-Hill, 1955.

[9] Goldman, L. W. "Principles of CT and CT Technology." *Journal of Nuclear Medicine Technology*, vol. 35, no. 3, 2007, pp. 115–128.

[10] Prince, Jerry L., Links, Johnathan M., *Medical Imaging Signals and Systems*. Prentice Hall, 2015.

[11] Bushberg, Jerrold T. *The Essential Physics of Medical Imaging, 3rd Ed*. Lippincott, Williams & Wilkins, 2012.

[12] Hansen, Per Christan. "The Truncated SVD as a Method for Regularization." *Numerical Analysis Project Manuscript NA-86-36*, Oct. 1986.

[13] Yang, L, et al. "AIR: Fused Analytical and Iterative Reconstruction Method for Computed Tomography." 5 Oct. 2013.

[14] Strohmer, Thomas, and Roman Vershynin. "A Randomized Kaczmarz Algorithm with Exponential Convergence." *Journal of Fourier Analysis and Applications*, vol. 15, no. 2, 2008, pp. 262–278.

[15] Candes, E.j., et al. "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information." ***IEEE Transactions on Information Theory*****, vol. 52, no. 2, 2006, pp. 489–509.**

[16] Candes, E.j., and M.b. Wakin. "An Introduction To Compressive Sampling." ***IEEE Signal Processing Magazine*****, vol. 25, no. 2, 2008, pp. 21–30.**

[17] Donoho, D.l., and X. Huo. "Uncertainty Principles and Ideal Atomic Decomposition." ***IEEE Transactions on Information Theory*****, vol. 47, no. 7, 2001, pp. 2845–2862.**

[18] Donoho, D.l., et al. "Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise." ***IEEE Transactions on Information Theory*****, vol. 52, no. 1, 2006, pp. 6–18.**

[19] Hanson, Kenneth M. "On the Optimality of the Filtered Backprojection Algorithm." ***Journal of Computer Assisted Tomography*****, vol. 4, no. 3, 1980, pp. 361–363.**

[20] Wurfl, Tobias, et al. "Deep Learning Computed Tomography: Learning Projection-Domain Weights from Image Domain in Limited Angle Problems." ***IEEE Transactions on Medical Imaging Special Issue on Machine Learning for Image Reconstruction*****, 2017.**

[21] Goodfellow, Ian, et al. ***Deep Learning*****. The MIT Press, 2017.**

[22] Strang, Gilbert. ***Linear Algebra and Learning from Data*****. Wellesley-Cambridge Press, 2019.**

[23] Nair, Vinod, and Geoffrey E Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines." *Proceedings of the 27 Th International Conference on Machine Learning*, 2010.

[24] Shu, Zhixin, et al. "Deforming Autoencoders: Unsupervised Disentangling of Shape and Appearance." *Computer Vision – ECCV 2018 Lecture Notes in Computer Science*, 2018, pp. 664–680.

[25] Ioffe, Sergey, and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." 2015.

[26] Mu, Yadong, et al. "Stochastic Gradient Made Stable: A Manifold Propagation Approach for Large-Scale Optimization." *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, 2017, pp. 458–471.

[27] Santurkar, Shibani. "How Does Batch Normalization Help Optimization?" *32nd Conference on Neural Information Processing Systems*, 2018.

[28] Gao, Hao. "Fused Analytical and Iterative Reconstruction (AIR) via Modified Proximal Forward–Backward Splitting: a FDK-Based Iterative Image Reconstruction Example for CBCT." *Physics in Medicine and Biology*, vol. 61, no. 19, 2016, pp. 7187–7204.

[29] Parikh, Neal, et al. *Proximal Algorithms*. Now Foundations and Trends., 2013.

[30] Polson, Nicholas G., et al. "Proximal Algorithms in Statistics and Machine Learning." *Statistical Science*, vol. 30, no. 4, 2015, pp. 559–581.

[31] Chen, Gaoyu. "AirNET: Fused Analytical and Iterative Reconstruction with Densely Connected Deep Neural Networks for Sparse-Data CT." ***Transactions on Medical Imaging*****, 2020.**

[32] Webb, Andrew R. ***Introduction to Biomedical Imaging*****. John Wiley & Sons, 2006.**

[33] Gonzalez, Rafael C., and Richard E. Woods. ***Digital Image Processing*****. Pearson, 2018.**

[34] Wang, Z., et al. "Image Quality Assessment: From Error Visibility to Structural Similarity." ***IEEE Transactions on Image Processing*****, vol. 13, no. 4, 2004, pp. 600–612.**

[35] Fleckenstein, Peter, et al. ***Anatomy in Diagnostic Imaging*****. Wiley-Blackwell, 2014.**

[36] Szegedy, Christian, et al. "Going Deeper with Convolutions." ***2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*****, 2015.**

[37] Shen, Liyue, et al. "Patient-Specific Reconstruction of Volumetric Computed Tomography Images from a Single Projection View via Deep Learning." ***Nature Biomedical Engineering*****, vol. 3, no. 11, 2019, pp. 880–888.**

[38] Li, Yinsheng, et al. "Learning to Reconstruct Computed Tomography Images Directly From Sinogram Data Under A Variety of Data Acquisition Conditions." ***IEEE Transactions on Medical Imaging*****, vol. 38, no. 10, 2019, pp. 2469–2481.**

[39] Gupta, Harshit, et al. "CNN-Based Projected Gradient Descent for Consistent CT Image Reconstruction." *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, 2018, pp. 1440–1453.

[40] Xie, Shipeng, et al. "Artifact Removal Using Improved GoogLeNet for Sparse-View CT Reconstruction." *Scientific Reports*, vol. 8, no. 1, 30 Apr. 2018.

[41] Mardani, Morteza, et al. "Recurrent Generative Residual Networks for Proximal Learning and Automated Compressive Image Recovery." 27 Nov. 2017.

[42] Yang, L, et al. "AIR: Fused Analytical and Iterative Reconstruction Method for Computed Tomography." 5 Oct. 2013.